

# TESTING THE ISOTROPY OF HIGH ENERGY COSMIC RAYS USING SPHERICAL NEEDLETS

BY GILLES FAÿ\*, JACQUES DELABROUILLE<sup>†,‡</sup>, GÉRARD  
KERKYACHARIAN<sup>‡</sup> AND DOMINIQUE PICARD<sup>‡</sup>

*Ecole Centrale Paris* \*, *CNRS* <sup>†</sup> and *Université Paris Diderot* <sup>‡</sup>

Ultra-high energy charged particles of unknown origin, which interact in the high atmosphere of the Earth generating large cascades of secondary particles, can be observed from the ground. For many decades, they have been a puzzle for particle physicists and astrophysicists. They seem to arrive from random directions in the sky, although the most energetic ones are supposed to point towards their sources with good accuracy. As an attempt to discriminate among several possible production scenarios, astrophysicists try to test the statistical isotropy of the directions of arrival of these cosmic rays. At the highest energies however, the observed cosmic rays are very rare, and testing the distribution of such small samples of directional data on the sphere is non trivial. We address here a nonparametric detection problem, making use of a multiscale analysis of the observation sample, based on a decomposition of the directional data using a wavelet frame on the sphere, the needlets. The aim is to test the isotropy, or the equality of the distribution with a known one. We explore two particular procedures, a multiple test and a plug-in approach. We describe the practical implementation of these two procedures and compare them to other methods in the literature. As alternatives to isotropy, we consider both very simple toy-models, and more realistic non isotropic models based on Physics-inspired simulations. The Monte Carlo study shows the good performances of the multiple test at moderate sample size, together with the robustness of its sensitivity with respect to the unknown characteristics of the alternative hypothesis. On the 69 most energetic events published by the Pierre Auger collaboration, our procedure detects significant departure from isotropy. The flexibility of this method and the possibility to modify it to take into account a large variety of extensions of the problem make it an interesting option for future investigation of the origin of ultra-high energy cosmic rays.

## 1. Introduction.

1.1. *Motivations.* It is a common problem in astrophysics to analyse data sets containing measurements of a number of objects (such as galaxies of a

---

*AMS 2000 subject classifications:* Primary 62G10, Nonparametric test, isotropy test, multiple tests, ultra-high energy cosmic rays, wavelet procedure; secondary 62C20, 62P35

*Keywords and phrases:* sample,  $\text{\LaTeX} 2\epsilon$

particular type) or of events (such as cosmic rays or gamma ray bursts) distributed on the celestial sphere. Each set of such objects or events can be represented as a collection of positions  $X_i = (\theta_i, \phi_i)$ ,  $i = 1, \dots, n$  in  $\mathbb{S}$  the unit sphere of  $\mathbb{R}^3$ . In many cases, such objects trace an underlying probability distribution  $f$  on the sphere, which itself depends on the physics which governs the production of the objects and events. Galaxies, for instance, form in over-densities of a preexisting smooth field of distribution of matter in the universe, and the study of the statistics of their distribution has grown into a field of astrophysics by itself (Martínez and Saar, 2002).

The case of ultra high energy cosmic rays (UHECRs) is of particular interest, and is the main focus of the present work. UHECRs are particles of unknown origin which arrive at the Earth from apparently random directions of the sky. These particles interact with atoms of the upper atmosphere, generating a huge cascade of billions of secondary particles. The observation of these secondary particles with appropriate detectors on ground permits the measurement of the direction of arrival and of the energy of the original cosmic ray.

The existence of cosmic rays has been known for about a century. Such particles exist with a very wide range of kinetic energies, from few eV to more than  $10^{20}$  eV.<sup>1</sup> Observed cosmic rays are typically ordinary charged particles (electrons, protons and nuclei), propagating in empty space, and deflected by galactic magnetic fields. The rate of observed cosmic rays in the vicinity of the Earth, however, decreases rapidly with energy. At low energy, the observed cosmic rays are numerous and their composition is well known. There also exist several known astrophysical processes responsible for their acceleration, such as stellar winds for the least energetic ones, to violent phenomena such as supernovae shock waves at higher energy. At the highest energies ( $E \geq 10^{20}$  eV), however, the observed flux is of the order of 1 event per square kilometre per century, which limits the statistics of observed events to few tens of events (in two decades of observations). In addition, no understood astrophysical process, involving known objects, can accelerate particles to such tremendous energies.

Recent observations of ultrahigh-energy cosmic rays suggest that they are ordinary particles, such as protons and nuclei, accelerated in extremely violent astrophysical phenomena (see ?, for a recent review on the astrophysics of UHECRs). However, many alternate hypotheses concerning their nature and origin have been proposed over the years (see, e.g., Hillas, 1984; Torres and Anchordoqui, 2004; ?). UHECRs could originate from active galactic nuclei (AGN), or from neutron stars surrounded by extremely high mag-

---

<sup>1</sup>1 eV = 1 electron Volt  $\simeq 1.6 \times 10^{-19}$  Joule

netic fields, or yet from many other processes. It is also possible that the type and origin of ultra high energy cosmic rays (at energies above  $10^{19}$  eV) depend, at least to some extent, upon the energy at which they are observed. Indeed, the most energetic cosmic rays cannot propagate very far (i.e. not much more than  $\sim 100$  Mpc), without losing most of their energy by interactions with photons from the Cosmic Microwave Background (the so-called GZK effect; [Greisen, 1966](#); [Zatsepin and Kuz'min, 1966](#)). The confirmation of the energy cut-off at the high end of the cosmic ray spectrum is one of the main achievements of the Pierre Auger Observatory ([Abraham et al., 2008](#); ?).

Before the location and physical process of acceleration have been clearly identified, taking into account the fact that most of the evidence about the chemical composition of cosmic rays at the highest energies rely on extrapolations of the present knowledge of hadronic interactions at energies two orders of magnitude above the range presently tested at the LHC, it is difficult to completely rule-out alternate theoretical explanations as to what UHECRs exactly are and what is their origin. Alternate hypotheses such as production by decay of long-lived relic particles from the Big Bang, about 13 billion years old ([Bhattacharjee and Sigl, 2000](#)), are just starting to be disfavored by the observations of the Auger collaboration, with recently published results about primary photon limits that impose stringent limits on these kind of models (?).

In an attempt to better understand the origin of such UHECRs, physicists study the statistical distribution of their directions of arrival, looking for two particular signatures. First, the (statistically significant) arrival of more than one UHECR from the same direction on the sky would indicate that their production is not likely to originate from single time events (e.g. catastrophic mergers of two compact astrophysical objects), but rather from sources which emit UHECRs regularly.<sup>2</sup> Second, one may look for correlation in the directions of arrival of UHECRs with known astrophysical objects, as nearby active galactic nuclei, in an attempt to identify plausible production sites. Hence, in some hypotheses, the underlying probability distribution for the directions of incidences of observed UHECRs would be a finite sum of point-like sources – or nearly point-like, taking into account the deflection of the cosmic rays by magnetic fields. In other hypotheses, the distribution could be uniform, or smooth and correlated with the local distribution of matter in the universe. The distribution could also be a superposition of the

---

<sup>2</sup>With the caveat that the time of propagation may depend on the energy and on the exact trajectory followed by the UHECR to reach us, making it possible that two particles reaching the Earth at different times have actually been emitted simultaneously.

above. Distinguishing between these hypotheses is of primordial importance for understanding the origin and mechanism of production of UHECRs.

In the past 20 years, a number of experiments have gathered observations of UHECRs, and several papers have been written which look for such features in the distribution of their directions of arrival, with sometimes contradictory conclusions. The difficulty lies in the fact that UHECRs are rare, and that they do not arrive necessarily exactly from the direction where their source is located. Indeed, as typical cosmic ray particles are charged (which permits their acceleration by electromagnetic processes), they are deflected by Galactic and intergalactic magnetic fields. The deflection depends on the length of the path through the magnetic field, and on the energy and charge of the particle. In fact, only very energetic cosmic rays (above few  $10^{19}$  eV) with small charge (e.g. protons or nuclei with small atomic numbers) are expected to travel typical astrophysical distances from their source to us with deflection angles smaller than a few degrees. Details of the deflections are not known, as neither the exact magnitude, orientation, and regularity on large scales of Galactic and extragalactic magnetic fields, nor the distance of the sources of UHECRs, nor the exact energy of the incoming cosmic ray, nor its charge (to within a factor of 26 between protons and iron nuclei), are known. Errors on the direction of the source of an UHECR can then be of order  $1^\circ$  at the lowest (typical error on the measurement of the direction of arrival with Auger), up to few degrees for protons, or tens of degrees for heavy nuclei travelling a long path through a regular galactic magnetic field.

Given a set of observed UHECRs, how can one best test for “repeaters” (cosmic rays coming from the same source) or more generally anisotropy in the distribution? If one restricts the analysis to the few events for which one is sure that the deflection angle is negligible, events are scarce and there is not enough statistics to conclude. As one selects events with less energy, the direction of origin becomes less reliable, with the total number of events completely dominated by those events with poorly constrained direction of origin. Finally, it is not clear how to build the isotropy test, without any sound prior knowledge about the uncertainty in the measured direction of the source. All of these are very meaningful questions to analyse UHECR observations.

Recently, an analysis of the direction of arrival of 27 UHECRs observed by the Pierre Auger experiment concludes in the existence of an anisotropy, and a correlation with objects in a catalogue of nearby active galactic nuclei (AGNs), located at distances lower than about 70 Mpc<sup>3</sup> ([Abraham et al., 2008](#)). This anisotropy, however, is less obvious in a more recent analysis,

---

<sup>3</sup>70 million parsecs  $\simeq 2.15 \times 10^{21}$  km

based on 69 observed events ([The Pierre AUGER Collaboration et al., 2010](#)). Clearly, the statistics is limited, and the development of new methods for investigating this topic can provide new insights on the origin of the UHECRs. Methods independent of external data sets such as the fore-mentioned VCV catalogue (which is not a statistically well-characterized sample of AGNs but a compilation of published results) are of particular interest.

*1.2. Outline of this work.* This work focuses on the important question of the isotropy of the cosmic rays. Because of the small number of available data, this question is not answered yet, although data from the Pierre Auger collaboration seems to hint at a correlation between the directions to the ultra-high energetic events (above  $5.5 \times 10^{19}$  eV) and the directions to active galactic nuclei in the catalogue compiled by Véron & Cetty-Véron (see [The Pierre AUGER Collaboration et al., 2008, 2010](#)). From a statistical point of view, we address the question of testing the goodness-of-fit of the isotropy assumption to this small sample of directional data. The framework we choose is purely nonparametric as we do not want to favour any particular alternative hypothesis, and as we wish to be able to discover unexpected forms of anisotropy.

The paper is organized as follows. In Section 2, we present a simplified model of cosmic ray propagation which will be used in Monte Carlo simulations to test the method. In Section 3 we present the nonparametric framework. Then we describe our needlet based anisotropy tests in Section 4. In section 5, we present a Monte Carlo experiment that compares the power of the different test in terms of sensitivity and robustness with respect to the parameters of the methods. We apply our procedures to real data from the Pierre Auger collaboration in Section 6. We then conclude and give perspectives for future extensions of the present work. Appendix A in the on-line supplement is devoted to a short description of the type of wavelets we have used (the needlets) and the practical and numerical implementation of our methods.

**2. Simulating cosmic ray emission.** In our investigation of tools to analyse the distribution of UHECR events, we need a way to simulate a distribution of observed events, as a function of an underlying physical model. A complete Monte Carlo simulation of the physical processes of cosmic ray emission and propagation in the magnetic fields is beyond the scope of this paper, and too dependent on a number of physical assumptions for which there is little available knowledge. We decide to perform qualitatively relevant simulations using a simple, although physically representative, toy model of cosmic ray emission and propagation.

2.1. *Cosmic ray sources.* In one hypothesis ( $H_0$ ), we will assume that cosmic rays are emitted from a uniform distribution of many sources, i.e. their directions of arrival are independent of the energy, and uniformly distributed on the celestial sphere. In the alternate hypothesis ( $H_1$ ), we will assume that  $n$  cosmic rays originate from a small number  $n_s$  of sources, distributed uniformly in a spherical volume  $V$  of universe, of radius  $r_{\max} = 70$  Mpc. For  $n_s \gg n$ , the distribution of directions of origin will be close to uniform, and ( $H_1$ ) indistinguishable from ( $H_0$ ). For  $n \gg n_s$ , and  $n_s$  small, coincidences in the directions of arrival of the observed UHECRs will permit to identify easily the directions of the emitting sources. Our objective is to address the issue when  $n_s$  is comparable to the number of observed events  $n$ .

Simulations are performed as follows:

- We fix the number  $n_s$  of sources and distribute them uniformly in the volume  $V$ . We assume that all sources are physically identical, i.e. they emit cosmic rays with the same probability, and the same distribution in energy, the latter coinciding with the observed flux  $dN/dE$ .
- We fix the number  $n$  of observed cosmic rays, and draw at random their energies according to the distribution  $n(E) \propto E^{-\alpha}$ ,  $E \in [E_{\min}, E_{\max}]$ ,  $\alpha > 0$ .
- For each observed cosmic ray, we assign at random a corresponding emitting source, according to a probability density inversely proportional to the square of the distance  $D$  to the source (sources nearer produce a larger flux on Earth). This probability distribution can be modulated by the acceptance of the instrument for studying realistic test cases. For instance, [The Pierre AUGER Collaboration et al. \(2010\)](#) use 69 highest energy events for the search of correlations with astrophysical sources, selected by a cut in zenith angle of arrival ( $\theta_{\text{zenith}} \leq 60^\circ$ ). Assumed homogeneous time coverage in UT over the years of observation, the coverage is computed straightforwardly from simple geometrical considerations (see ?, and the details at the end of Section 2.2). The map of Auger coverage computed in this way is displayed in figure 1. The effect of the GZK cutoff is taken into account simply by limiting the volume to a sphere of 70 Mpc radius.
- For each cosmic ray, we modify the direction of arrival due to extragalactic magnetic fields. The next subsection describes the model used to implement these deflections.

2.2. *Deflection by Galactic and extragalactic magnetic fields.* Galactic magnetic fields are an important component of the Galactic interstellar

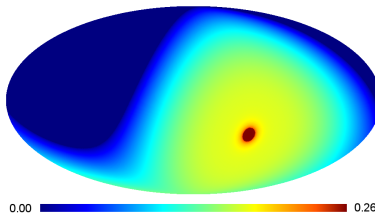


Fig 1: Coverage function  $g$  for the Pierre-Auger observatory in Galactic coordinates, represented through a Mollweide projection and computed from geometrical considerations (see ?). Moreover,  $g$  has been rescaled to be a probability density function on the sphere.

medium (ISM). They can be probed in a variety of ways. The impact of local magnetic fields is observed in the optical wavelength range via starlight polarization, as elongated interstellar dust grains in the foreground of the observed star, aligned perpendicularly to magnetic field lines, absorb preferentially one direction of starlight polarisation. Measurements of many stars reveal a general picture of the magnetic field in the Milky Way near the Sun (??). Aligned dust grains also emit polarized infrared emission, which can be used to infer magnetic fields in dust clouds (?). Zeeman splitting of radio spectral lines allows for the direct measurement of relatively strong fields in nearby, dense gas clouds in the Milky Way (?). On larger-scales, the magnetic field of our Galaxy can be probed in three dimensions using Faraday rotation of pulsar signals (?). Finally, synchrotron emission, emitted by relativistic electrons spiralling in the magnetic field, can be used to constrain the direction and amplitude of the magnetic field either from direct observation of the synchrotron polarisation (?), or by measuring the Faraday rotation of Galactic synchrotron using multi-wavelength observations in the radio range (below few GHz) (?).

In the vicinity of the Sun, the Galactic magnetic field has a typical amplitude of few microGauss. This amplitude is typically increasing with decreasing distance towards the Galactic center, where it can reach values of few tens of microGauss, and up to a few milliGauss in very local regions. In general the regular component over most of the outer Galaxy is of order few microGauss, aligned along the Galactic plane. The overall field structure follows the optical spiral arms, with evidence for at least one large-scale field reversal in the disk, inside the solar radius, and several distortions near star-forming regions.

For the purpose of estimating their impact on the deflection of high energy cosmic rays, Galactic magnetic fields are typically modeled as the sum of



two components with different physical properties, a regular component, and a turbulent component. The regular component roughly follows the spiral arms of the Galaxy, and induces deflections typically perpendicular to the Galactic plane, i.e. deflections in latitude of arrival. The turbulent component induces random deflections, which can be modeled as two-dimensional Gaussian distributions centered at the source. Indeed, we assume that such deflections are made of the superposition of many independent small deflections by independent regions with independent magnetic field directions, so that the Gaussian hypothesis is justified by the central limit theorem. We consider only cases in which the total deflection is small enough that the projection to the sphere is irrelevant (as well as the truncation of angles to  $2\pi$ ). Typical deflections for atomic nuclei are as follows (Harari et al., 2002).

For the regular component (magnetic lensing effect):

$$\delta_{\text{reg}} = 3.25^\circ \left( \frac{10^{20} \text{ eV}}{E/Z} \right) \left( \frac{B}{2 \mu\text{G}} \right) \left( \frac{r}{3 \text{ kpc}} \right) \quad (1)$$

where  $E$  is the energy of the UHECR in eV,  $Z$  is the atomic number (e.g. 1 for hydrogen nuclei (protons), 2 for Helium nuclei (alpha articles), etc.),  $B$  is the magnetic field in micro Gauss ( $\mu\text{G}$ ), and  $r$  the propagation length of the cosmic ray in the magnetic field. The deflection is assumed deterministic (although energy-dependent), and the instantaneous direction of the deflection is along  $\vec{v} \times \vec{B}$ , where  $\vec{v}$  is the velocity of the incoming particle and  $\vec{B}$  the regular Galactic magnetic field, assumed to be along the  $y$ -axis of the Galactic coordinate system.

For the turbulent component (random deflection):

$$\delta_{\text{turb}} = 0.56^\circ \left( \frac{10^{20} \text{ eV}}{E/Z} \right) \left( \frac{B}{4 \mu\text{G}} \right) \sqrt{\frac{r}{3 \text{ kpc}}} \sqrt{\frac{L_{\text{gal}}}{50 \text{ pc}}} \quad (2)$$

The deflection is Gaussian distributed with a standard deviation  $\delta_{\text{turb}}$ , and uniform distribution of the direction of the deviation in  $[0, 2\pi[$ . The deflections are written in terms of the typical expected values for the magnetic field ( $2 \mu\text{G}$  for the regular part and  $4 \mu\text{G}$  for the turbulent part), for coherence length  $L_{\text{gal}}$  of the turbulent part of the Galactic magnetic field (about 50 pc). 3 kpc is the typical propagation length  $r$  inside the Galactic magnetic field for a cosmic ray coming perpendicularly to the Galaxy. A plane parallel approximation of the disc-shaped geometry of the Milky Way suggests a dependence of  $r$  with the Galactic latitude  $b$  of the incoming cosmic ray. We assume here a dependence  $r \propto 1/\sin b$ , with a maximum length of 10 kpc, typical of the size of the Galactic disk.



Extragalactic magnetic fields also deflect cosmic rays originating from distant locations in the Universe. These deflections are expected to be qualitatively similar to those due to the turbulent part of the Galactic magnetic field, except that typical field strengths are smaller (and less well known) and correlation lengths are larger. Following The Pierre Auger collaboration ([The Pierre AUGER Collaboration et al., 2008](#)), we assume a deflection with standard deviation given by:

$$\delta_{\text{ext}} = 2.4^\circ \left( \frac{10^{20} \text{ eV}}{E/Z} \right) \left( \frac{B}{1 \text{ nG}} \right) \sqrt{\frac{D}{100 \text{ Mpc}}} \sqrt{\frac{L_{\text{ext}}}{50 \text{ pc}}} \quad (3)$$

UHECRs are observed to arrive on Earth with a flux  $dN/dE$  proportional to  $E^{-4.2}$  for energies  $E > 4 \times 10^{19}$  eV ([Abraham et al., 2008](#)). Although the shape of the spectrum is not very well constrained in this region (more recent Auger results suggest a spectral index closer to  $-4.3$ ), the exact shape of the spectrum does not have a strong impact on the validity of our analysis. Our simulations will assume such a distribution, with various values for the minimum energy  $E_{\text{min}}$ , and  $E_{\text{max}} = 10^{21}$  eV. We focus on very energetic UHECRs ( $E > 10^{19}$  eV), and assume UHECRs are light nuclei ( $Z \approx 1$ ), for which deflections by magnetic fields are expected to be of the order of a few degrees.

We then implement cosmic ray deflections according to equation (3) (first the cosmic ray first travels in the intergalactic medium), and then using both equations (1) and (2). As the exact nature of the cosmic rays has little impact on the general principles of our method, except that a change in atomic number induces a change in the scale of the deflections, we have assumed here for simplicity that all cosmic rays are protons (i.e.  $Z = 1$ ). This however, as a further refinement, can be easily changed for practical application on real data sets. In particular, the presence or lack of anisotropy in the directions of arrival of the highest energy cosmic rays may help shed light on the nature of these particles, as iron nuclei, for instance, are more deflected by magnetic fields than protons, by a factor  $Z_{\text{iron}} = 26$ . This is an important point to take into account in view of recent Auger results that seem to indicate a low proton fraction at energy above  $10^{18}$  eV, so that the cosmic rays at those energies might be essentially heavier nuclei (?).

Figure 2 illustrates simulated outcomes in two extreme cases: few sources and many cosmic rays (right) and many sources and few cosmic rays (left).

In practice, instruments observe the sky unevenly. The capability of the instrument to observe in a particular direction of the sky depends on the field of view of the instrument, and on the orientation of the instrument with respect to the sky (which itself depends on the sidereal time). From

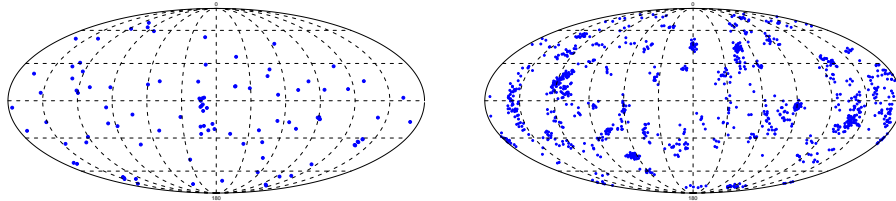


Fig 2: Two simulations of the physical model described in Section 2, with  $\alpha = 4.2$ ,  $E_{\min} = 4 \times 10^{19}$ ,  $E_{\max} = 10^{21}$ . On the left, the number of sources is  $n_s = 1000$  and the number of observations is  $n = 100$ . On the right,  $n_s = 100$  and  $n = 1000$ . It appears in this latter case that clusters of events are of different typical angular size.

the properties of the instrument and the geometry of the observations, one can infer an equivalent observing time as a function of direction on the sky, i.e. a function on the sphere that modulates the probability of detection of the observed cosmic rays. As an illustration, we have displayed on Figure 1 a Mollweide projection of the coverage map associated with the Pierre Auger observatory, in Galactic coordinates, computed following section 2 in ?. This coverage map has been generated assuming a maximum accepted zenith angle for incoming cosmic rays of  $\theta_{\text{zenith}} = 60$ , and uniform distribution of observation periods in universal time (and hence, a coverage that depends exclusively of the declination, not the right ascension, in equatorial coordinates). The effect of the precession of equinoxes has been neglected for generating this coverage map (the perturbations it would generate are very tiny as compared to what we can measure with about 100 events, as currently available).

### 3. Nonparametric tests on the sphere.

**3.1. Introduction.** Suppose that  $(X_1, \dots, X_n)$  is an  $n$  sample of i.i.d random positions on the two-dimensional sphere with probability density function  $f$ . Let  $f_0 \equiv 1/4\pi$  denote the uniform density on  $\mathbb{S}$ . Our aim, as explained in the previous section is to test

$$(H_0) : f \equiv f_0 \quad \text{against} \quad (H_1) : f \not\equiv f_0. \quad (\mathcal{T}_1)$$

One can take into consideration the non-uniform angular acceptance in the observation model by considering some known and arbitrary function  $g : \mathbb{S} \rightarrow [0, 1]$ . Incoming events from direction  $\xi \in \mathbb{S}$  have probability  $g(\xi)$  to be

observed by the instrument. In this case the observed directions  $X_1, \dots, X_n$  are distributed along a density which is proportional to  $g$  under the null hypothesis (isotropy of the original distribution). Let divide  $g$  by  $\int_{\mathbb{S}} g(\xi) d\xi$  so that  $g$  is now normalized to be a density probability function on the sphere. In order to test for isotropy of the physical phenomenon in this context, we need to implement the test

$$(H_0) : f \equiv g \quad \text{against} \quad (H_1) : f \not\equiv g. \quad (\mathcal{T}_2)$$

On the real line, testing for  $f \equiv g$  can be reformulated as testing for the uniform distribution of the sample  $G(X_1), \dots, G(X_n)$  on  $[0,1]$  where  $G$  is the distribution function associated with the probability density  $g$ . For higher dimensions (as on the sphere), there is no natural transformation of the data, no notion of distribution function for directional data, that allows to recast  $\mathcal{T}_2$  as  $\mathcal{T}_1$ . Then we consider  $\mathcal{T}_2$  in its generality, with  $\mathcal{T}_1$  as a particular case.

Our aim in this paper is to provide test algorithms which are at the same time easy to implement, efficient in practical situations where the sample size is small (a few tens) and the data may be collected in a non uniform nor complete way, but also with properties that are likely to be optimal from a theoretical point of view.

Let us begin with a short review on nonparametric tests associated to function estimation, since this will inspire our study in many ways.

**3.2. Anisotropy tests among general nonparametric tests.** The test problem is well posed when the alternative is given. More often in practice it is wiser to consider a large nonparametric class of alternatives. To allow derivation of optimality properties, following standard point of view in a nonparametric framework (see for instance [Ingster and Suslina, 2003](#); [Ingster, 1993](#); [Butucea and Tribouley, 2006](#)), we shall consider smooth alternatives of the form

$$(H_{1,n}) : f \in \mathcal{F}_n(d, C) \quad (4)$$

where

$$\mathcal{F}_n(d, C) = \{h \in \mathcal{R} : d(h, g) > Cr_n\} \quad (5)$$

and  $\mathcal{R}$  is a class of regularity, that contains for example all the twice continuously differentiable densities, or densities satisfying the Hölder condition with Hölder exponent  $s > 0$ . We may consider balls in Sobolev or Besov spaces (see below). Here,  $d$  is a (semi-) distance between densities and  $r_n$  is referred to as a *separation rate*. Roughly speaking,  $d$  and  $r_n$  respectively define the shape and the size of the neighbourhood of the density under the

null which is excluded from the alternative set of densities. The multiplicative constant  $C$  allows to define the concept of critical separation rate; see Eqs. (8) and (9) below.

The choice of such alternatives is essential for the test procedure because the test statistics are built, more or less, on estimators of  $d(f, g)$ . For some particular distances, nonparametric estimators  $\hat{f}$  of the density of the observed sample may be plugged into the distance, namely

$$\hat{d}(f, g) = d(\hat{f}, g).$$

For instance,  $\hat{f}$  could be an histogram-like (pixel-wise constant) density estimate of  $f$  based on counting events falling in any pixel of a given tessellation  $\{V_k\}_{k=1, \dots, K}$  of the sphere, namely

$$\hat{f} = \frac{1}{n} \sum_{k=1}^K \#\{X_i \in V_k, i = 1, \dots, n\} \frac{\mathbf{1}_{V_k}}{\mu(V_k)}$$

and the decision could be taken on the value of  $d(\hat{f}, g) = \|\hat{f} - g\|^2$ , say. Nevertheless, as described in Ingster (2000), such “plug-in” procedures are not always optimal in terms of rates of separation (see Section 4.2 for a more precise statement). In contrast, multiple tests have nice theoretical (minimax optimality and adaptivity) properties in various contexts: detection in a white noise model (Spokoiny, 1996),  $\chi^2$  test of uniformity on  $[0, 1]$  (Ingster, 2000), goodness-of-fit test and model selection for random variables on the real line (Fromont and Laurent, 2006), two-sample homogeneity tests (Bucea and Tribouley, 2006), for instance. Note that one would also like to test for uniformity by taking into account the uncertainty on the measurements of the directional data. In a first approximation, this error can be modeled as a convolution noise: the observations are  $Z_i = \epsilon_i X_i, i = 1, \dots, n$  where  $\epsilon_1, \dots, \epsilon_n$  is an i.i.d. sequence of random rotations in  $\text{SO}(3)$ . ? addressed the problem of testing for the isotropy  $(X_1, \dots, X_n)$  in the particular case of uniform coverage of the sphere and from noisy observation (random rotations of the directions). As a consequence of the uniform coverage, their adaptive testing procedure is ideally constructed on the multipole moments of the observations.

If one has strong prior information, it is possible to construct tests that are not uniform except along a few set of directions, but which can have as much power as possible at the  $n^{-1/2}$  scale in those few directions of interest. This framework is introduced in ? and applied in ? for detecting periodicity in a sequence of photon arrival times. We insist on the fact that our setting

is different as we do not assume any a priori knowledge on the alternative hypothesis.

In the following paragraphs we discuss the various ingredients of our study.

**3.2.1. Distances.** We will consider standard distances of functions on the sphere, although there is in fact no clear choice for a 'good' distance in this framework :  $L^1$  distance is generally more appropriate for probability densities, but  $L^p$  distances when  $p$  is increasing and especially  $L^\infty$  are more and more sensitive to bumps. As it is both usual and practical, we will mainly consider the  $L^2$  distance (with respect to the invariant measure on the sphere). But, we will also consider to express our results other  $L^p$  distances such as  $L^1$  and  $L^\infty$ . It is important to notice that it is the remarkable ability to concentrate of the needlets that enables us to consider various distances. More traditional bases would only allow the  $L^2$  distance and would then be much less sensitive to local changes.

**3.2.2. Separation rate.** Let  $T(X_1, \dots, X_n) \in \{0, 1\}$  be a non randomized decision, *i.e.* a measurable function of the sample  $(X_1, \dots, X_n)$  with value in  $\{0, 1\}$ . The dependence in  $n$  is omitted in most of our notations. As usual the event  $[T = 1]$  is equivalent to the rejection of the null hypothesis. The probability of error of the first kind (false positive) of the decision is denoted

$$\alpha_n(T) = \mathbb{P}_g(T = 1) \quad (6)$$

while probability of error of the second kind (false negative) against the alternative (4) is

$$\beta_n(T, C) = \sup_{f \in \mathcal{F}_n(d, C)} \mathbb{P}_f(T = 0). \quad (7)$$

Here  $\mathbb{P}_f, \mathbb{P}_g$  denote the probability measure under the density  $f$  or  $g$  for the i.i.d sample  $(X_1, \dots, X_n)$ .

Formally the separation rate is defined using the following minimax optimality criterion. A sequence  $r_n$  is a minimax rate of testing (see [Ingster, 2000](#)) if the following statements are satisfied:

1. For any  $r'_n$  such that  $r'_n/r_n \rightarrow 0$  as  $n \rightarrow \infty$ ,

$$\liminf_{n \rightarrow \infty} \inf_T \{\alpha_n(T) + \beta_n(T, 1)\} = 1 \quad (8)$$

where the infimum is taken on all decision rules, *i.e.*  $\{0, 1\}$ -valued measurable functions of the sample  $(X_1, \dots, X_n)$ .

2. For any  $\alpha, \beta > 0$ , there exist some constant  $C > 0$  and a test statistic  $T^*$  (said *rate optimal in the minimax sense*), such that

$$\limsup_{n \rightarrow \infty} \alpha_n(T^*) \leq \alpha \quad \text{and} \quad \limsup_{n \rightarrow \infty} \beta_n(T^*, C) \leq \beta \quad (9)$$

Condition (8) says that if the separation rate vanishes faster than  $r_n$ , then no test can do better than the blind random decision, for which the sum of the errors of the two kinds is exactly 1. Condition (9) says that there exist a decision that is efficient for such a separation rate, so that this rate is indeed a critical rate.

It is clear that a good test become sensitive to closer and closer alternative hypothesis ( $H_{1,n}$ ) when the sample size  $n$  grows. The critical radius gives a precise and quantitative for this behavior. The rate  $r_n = 1/\sqrt{n}$  is the usual rate in the regular parametric setting.

**3.2.3. Invariance properties.** As the uniform distribution is invariant under rotations of the sphere, the theory of invariant tests (see [Lehmann and Romano, 2005](#), chapter 6) leads to impose the same kind of invariance on any statistical procedure for deciding whether  $f = f_0$  or not (see for instance [Giné \(1975\)](#) and the references therein). As bases of invariant subspaces under rotations, the spherical harmonics are thus the most natural tools to detect some deviation from isotropy as in problem  $(\mathcal{T}_1)$ . However, as explained earlier, a common property of astrophysical observation of (point or continuous) processes on the sphere is the non-uniform coverage of the *sky* by the instrument. It is common also that some parts of the data are missing or so noisy that it is preferable to completely ignore or mask them. That is why non invariant approaches must be considered, and localized analysis functions (such as wavelets) may be used as alternatives to spherical harmonics. In the same spirit, wavelets has been proposed in the context of the angular power spectrum estimation by [Baldi et al. \(2008\)](#) and used in the realistic case of a partially observed stationary process with heteroscedastic noise in [Faÿ et al. \(2008\)](#) and [Faÿ and Guillaux \(2011\)](#).

**3.2.4. Regularity conditions: Besov spaces on the sphere.** Although this is not directly the purpose of this paper, it is a natural question to ask which kind of regularity spaces our procedures are designed for. The problem of choosing appropriate spaces of regularity on the sphere is a serious question, and it is important to consider the spaces which generalize usual approximation properties. On the other hand, we are interested in spaces of functions which can be characterised by their needlet coefficients  $\{\beta_{jk}\}$  associated to a needlet frame  $\{\psi_{jk}\}$  (where  $j$  denote the scale and  $k$  the

position; see Appendix A in the online supplement for the precise definitions). Hence, as is standard in the nonparametric literature, it is natural to consider Besov bodies constructed on the needlet basis. In many situation (not only the sphere) it can be proved that these spaces can also be described as approximation spaces, so they have a genuine meaning and can be compared to Sobolev spaces. We define here the Besov body  $B_{pq}^s$  as the space of functions  $f = (4\pi)^{-1} \int_{\mathbb{S}} f d\mu + \sum_{j \geq 0} \sum_{k \in \mathcal{K}_j} \beta_{j,k} \psi_{j,k}$  such that

$$\sum_{j \geq 0} 2^{jsq} \left( \sum_{k \in \mathcal{K}_j} (|\beta_{j,k}| \|\psi_{j,k}\|_p)^p \right)^{q/p} < \infty$$

(with the obvious modifications for the cases  $p$  or  $q = \infty$ ). Details on Besov spaces and their characterization by wavelets can be found in [Triebel \(1992\)](#) and [Meyer \(1992\)](#). For details on the relations between needlets and Besov spaces we refer for instance to [Narcowich et al. \(2006a,b\)](#), [Petrushev and Xu \(2008\)](#) and [Kerkycharian and Picard \(2009\)](#).

**4. Needlet based test procedure and other anisotropy tests.** We introduce here two anisotropy detection procedures based on the needlet analysis of  $\{X_i\}_{i=1,\dots,n}$ . The first one is based on multiple testing and will be referred to as MULTIPLE. The second one uses an estimate of the density plugged in a distance criterion, and will be referred to as PLUGIN. For sake of further comparison (see Section 5), we also describe two existing methods that are used in the gamma ray burst and cosmic ray literature. The first one is based on a nearest neighbour analysis (see [Quashnock and Lamb, 1993](#); [?; Efron and Petrosian, 1995](#)). The second one relies on the two-point correlation (see e.g. [Kachelriess and Semikoz, 2006](#)).

We want detection procedures that are efficient from a  $L^2$  point of view, but also for other  $L^p$  norms. In addition, we will require procedures that are simple to implement as well as adaptive to unknown and inhomogeneous smoothness. In Euclidean frameworks, these types of requirements are well-known to be efficiently handled by (non-linear) wavelet thresholding estimation in the context of density estimation (see for instance [?](#)), or by multiple tests ([Ingster, 2000](#); [Spokoiny, 1996](#)).

Our problem here requires a special construction adapted to the sphere, since usual tensorized wavelets will never reflect the manifold structure of the sphere and will necessarily create unwanted artifacts. Recently a tight frame (i.e. a redundant family sharing some properties with orthonormal bases), called *needlet frame*, was produced which enjoys enough properties to be successfully used for density estimation ([Baldi et al., 2009](#)) e.g. concentration in the “Fourier” domain as well as in the space domain. Here, obviously the



“space” domain is the two-dimensional sphere itself whereas the Fourier domain is now obtained by replacing the “Fourier” basis by the basis of Spherical Harmonics which leads as mentioned in the previous section to invariant tests. This construction produces a family of functions  $\{\psi_{jk}, j \geq 0, k \in \mathcal{K}_j\}$  which very much resemble wavelets. The index  $k$  defines (with an analogy to the standard wavelets) the *locations* (points) on the sphere around which the needlet is concentrated, and  $j$  is referred to as the *scale*. These needlets have been shown to be extremely useful for solving several type of astrophysical problems (Delabrouille et al., 2009; Faÿ et al., 2008; Pietrobon et al., 2006; Marinucci et al., 2008; Pietrobon et al., 2008; Rudjord et al., 2009) or diverse inverse problems in statistics (Kerkycharian et al., 2007, 2011, 2010). They are especially well adapted to the situation recurrent in astrophysics, where the “full sky” is not covered (meaning in our context that there are regions of the sphere where the points  $X_i$  are not observed if they happen to fall there).

A formal definition of needlets on the sphere is postponed to Appendix A (on-line supplement) and can be found in greater details in Narcowich et al. (2006b). For the description of the test procedures, we only need to define the empirical needlet coefficients

$$\hat{\beta}_{jk} \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \psi_{jk}(X_i) \quad (10)$$

which are unbiased estimators of  $\beta_{jk}(f) \stackrel{\text{def.}}{=} \langle f, \psi_{jk} \rangle = \int_{\mathbb{S}} f(\xi) \psi_{jk}(\xi) d\xi$ . As usual in the wavelet literature,  $j \geq 0$  refers to the scale and  $k$  to the location. The coarsest scale is  $j = 0$ . The index  $k$  refers to a collections of quadrature points  $\{\xi_{j,k}\}$  that are available at each scale  $j$ .  $\psi_{j,k}$  is then a zero-mean function centered on  $\xi_{j,k}$  and more and more concentrated as  $j \rightarrow \infty$ .

**4.1. Multiple tests.** For multiple tests, we will consider collections of “linear estimators” of the density, meaning that we won’t use any non-linear processing of wavelet coefficient such as thresholding in the estimation phase. By analogy with the work of Butucea and Tribouley (2006) on the related problem of the two-sample nonparametric homogeneity test, we define

$$\hat{f}_J = \frac{1}{4\pi} + \sum_{j=0}^J \sum_{k \in \mathcal{K}_j} \hat{\beta}_{jk} \psi_{jk} \quad (11)$$

with the  $\beta_{jk}$ 's given by (10). For any value of the smoothing parameter  $J$ , we define the non-randomized associated testing procedure

$$T_J = \mathbf{1}_{d(\hat{f}_J, g) \geq t_J} = \begin{cases} 1 & \text{if } d(\hat{f}_J, g) \geq t_J \\ 0 & \text{if } d(\hat{f}_J, g) < t_J \end{cases} \quad (12)$$

This gives a family of tests indexed by  $J$ , where the dependence with respect to the choice of the distance  $d$  and to the sequence of thresholds  $t_J$  is made implicit in the notation.

Butucea and Tribouley (2006) proved that if the regularity conditions are known and specified by Besov conditions, the smoothing parameter  $J$  can be chosen optimally. It is likely that their arguments could be reproduced in our case. However our point of view in this paper will not be to detail this theoretical issue but rather to concentrate on the practical aspects of the tests. Moreover, it would be probably difficult to relate physical information to mathematical regularity conditions.

Nevertheless, the optimal choice for the parameter  $J$  depends on the regularity  $s$  specified in the class of alternatives. Adaptive optimality can be achieved thanks to a multiple test that decides for the alternative hypothesis as soon as one of the  $T_J(d, c_J) = 1$  individually does so, *i.e.* defining  $T^{\text{MULTIPLE}}$ , by

$$T^{\text{MULTIPLE}} = 0 \text{ if and only if } \forall J \leq J^*, T_J = 0. \quad (13)$$

Mimicking the theoretical results obtained in Butucea and Tribouley (2006) and Baldi et al. (2009) we have used

$$J^* = \lfloor \frac{1}{2} \log_2(n / \log n) \rfloor \quad (14)$$

as reference in our numerical investigations, as in the case of adaptive density estimation (see below). Note however that the optimal  $J^*$  could vary according to the loss function ( $L^p$  norm) we use to measure the non isotropy as suggested by the results of the related problem in the two sample nonhomogeneity detection Butucea and Tribouley (2006). The values  $t_J$  that are used in (12) must be chosen to verify  $\mathbb{P}_g(T^{\text{MULTIPLE}} = 0) \simeq \alpha$  where  $\alpha$  is the prescribed level of the test.

**4.2. Plug-in tests.** It is also interesting to compare, from an empirical point of view, the above multiple test procedures to algorithms where we simply plug-in an adaptive estimate of the density in the distance. These density estimators have good asymptotic properties from a minimax point of view, hence it makes sense to investigate also their properties when used

for testing. To the best of our knowledge no theoretical optimality is proved and there even are arguments suggesting that these procedures might not be optimal. For instance, on the real line, the minimax rate of convergence for estimation (in the so-called dense case) is  $n^{-s/(2s+1)}$ , meaning that if  $f$  belongs to a ball in a Hölder space with exponent  $s$ , then no estimator can approach the least favorable density at a better error rate (measured in a  $L^p$  norm). We refer to ?, Theorem 3 for a precise statement, among others. On the other hand, the minimax critical radius for non-uniformity detection is  $n^{-2s/(4s+1)}$  (see [Ingster, 2000](#)). It means that, in the minimax framework, one can distinguish asymptotically two hypotheses that are separated by a distance negligible with respect to the accuracy of any nonparametric estimation of the densities in a infinite dimensional space.

The most popular minimax adaptive technique consist in adding to a very basic linear estimation a thresholding rule as post-processing. In the above mentioned paper ([Baldi et al., 2009](#)) this nonlinear post-processing actually is a *hard* thresholding rule, namely

$$\hat{f}_{J^*} = \frac{1}{4\pi} + \sum_{j=0}^{J^*} \sum_{k \in \mathcal{K}_j} \hat{\beta}_{jk} \mathbf{1}_{|\hat{\beta}_{jk}| > \kappa \sqrt{\log n/n}} \psi_{jk}$$

for some positive constants  $\kappa$  and  $J^* = \lfloor \frac{1}{2} \log_2(n/\log n) \rfloor$ . The coefficients  $\hat{\beta}_{jk}$  are defined in (10).

It is known that many variations exist with close theoretical properties but some differences in different practical situations. Among those, we will especially consider the data-driven thresholding introduced by [Juditsky and Lambert-Lacroix \(2004\)](#) to deal with density estimation on the real line (as opposed to density on  $[0,1]$ ). It seems to give good detection procedures for small samples in our context. In the following, we will consider the non-linear estimates

$$\hat{f}_{J^*} = \frac{1}{4\pi} + \sum_{j=1}^{J^*} \sum_{k \in \mathcal{K}_j} \mathbf{1}_{|\hat{\beta}_{jk}| > \lambda \sqrt{\log n} \hat{\sigma}_{jk}} \mathbf{1}_{\delta_{jk} > \rho \log n} \hat{\beta}_{jk} \psi_{jk} \quad (15)$$

for some positive constants  $\rho, \lambda$ ,  $J^* = \lfloor \frac{1}{2} \log_2(n/\rho \log n) \rfloor$ , and where

$$\hat{\sigma}_{jk}^2 \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \psi_{jk}^2(X_i) - (\hat{\beta}_{jk})^2, \quad (16)$$

$$\delta_{jk} \stackrel{\text{def.}}{=} (\psi_{jk}^2(\xi_{jk}))^{-1} \sum_{i=1}^n \psi_{jk}^2(X_i). \quad (17)$$

Let us give a short interpretation of the thresholding procedure. The quantity  $\hat{\sigma}_{jk}^2$  is an estimate of the variance of  $\hat{\beta}_{jk}$ . The expression for  $\delta_{jk}$  is inspired by the one provided in [Juditsky and Lambert-Lacroix \(2004\)](#). In this reference, compactly supported wavelets on the real line are used with a threshold on the number of observations actually participating to the estimation of  $\beta_{jk}$ . In this case, it makes sense to count the number of observations falling in the support of the wavelet. In our case, as needlets are supported on the whole sphere (although very concentrated), we propose to replace this quantity by a continuous type approximation  $\delta_{jk}$ , see (17). Note that  $\delta_{jk} = n$  if  $X_1 = \dots = X_n = \xi_{jk}$ .

Finally, we define the PLUGIN procedure as the decision

$$T_J^{\text{PLUGIN}} = \mathbf{1}_{d(\hat{f}_{J^*}, g) \geq t_J^{\text{PLUGIN}}} \quad (18)$$

with  $\hat{f}_{J^*}$  defined in (15) and  $t_J^{\text{PLUGIN}}$  some fixed threshold depending on the prescribed level  $\alpha$  of the test.

*4.3. Two-point correlation test and nearest neighbour test.* When dealing with one dimensional data, one can compare every test procedure to the well known benchmark Kolmogorov-Smirnov or Cramér-von Mises tests, which are based on the empirical distribution function of the sample. In higher dimensions (here on the sphere), there is no natural order relation that allows to consider such approaches. For sake of comparison, we have run some simulations on two different tests found in the astronomical literature.

*Nearest neighbour test.* The following statistical procedure has been proposed by Quashnock & Lamb (1993). We denote it NN, as nearest neighbour. For each point  $X_i$ , we compute the distance  $Y_i$  to its nearest neighbour. Under the hypothesis that  $f$  is uniform over the whole sphere, the marginal distribution function of  $(Y_i)$  is  $\phi : y \mapsto 1 - [(1 + \cos y)/2]^{n-1}$ , and the Wilcoxon statistic

$$W = \sqrt{12n} \left( \frac{1}{2} - \frac{1}{n} \sum_{i=1}^n \phi(Y_i) \right)$$

is asymptotically standard Gaussian. For nonhomogeneous random draw (for instance, in presence of clusters), this statistic is expected to take significantly high values, allowing to detect this kind of anisotropy. This test is of interest as it is simple to compute, it has no parameters to be tuned, and that it admits an extension to non uniform sky coverage (see [Efron and](#)

[Petrosian, 1995](#)). In this case, the distribution of  $W$  is estimated numerically by Monte Carlo methods. The NN procedure simply writes

$$T^{\text{NN}} = \mathbf{1}_{W \geq t^{\text{NN}}} . \quad (19)$$

where  $t^{\text{NN}}_{1-\alpha}$  is the  $(1-\alpha)$ -quantile of the distribution of  $W$ . This distribution can be approximated by a standard Gaussian distribution if the sample size is big and the coverage is uniform. Otherwise, the quantile is estimated by Monte Carlo method.

*Two-point correlation test.* Among others, [Kachelriess and Semikoz \(2006\)](#) use the empirical two-point auto-correlation function to detect clustering (TwoPC test). For a collection of  $n$  points  $\{X_i\}$  and any angular distance  $\delta \in [0, \pi]$ , let  $N_n(\delta)$  denote the random number of pairs  $\{i, j\}$  such that  $\Delta(X_i, X_j) \leq \delta$ , where  $\Delta$  is the geodesic distance. Define the two-point correlation function  $w_n(\delta) = \mathbb{E}(N_n(\delta))$  and its empirical counterpart

$$\hat{w}_n(\delta) = \sum_{i < j} \mathbf{1}_{[0, \delta]}(\Delta(X_i, X_j)) . \quad (20)$$

Under the null hypothesis, the distribution of  $\hat{w}_n$  at any  $\delta_0$  is evaluated using Monte Carlo simulations. Then, the detection will be based on the comparison between the empirical correlation function and  $w_n$ , at some fixed value  $\delta_0$  or a few different values. A typical  $\delta_0$  can be chosen so as to maximize the sensitivity of the test depending on the application. In some references however the probability to observe a value bigger than  $\hat{w}_n(\delta)$  is plotted on the whole range  $[0, \pi]$  with no  $\delta_0$  fixed a priori. Consequently, much care be taken when interpreting those values, as stressed for instance in [Kachelriess and Semikoz \(2006\)](#). Here we define the procedure TwoPC by the decision

$$T^{\text{TwoPC}} = \mathbf{1}_{\hat{w}_n(\delta_0) \geq t^{\text{TwoPC}}} . \quad (21)$$

where  $t^{\text{TwoPC}}_{1-\alpha}$  is the  $(1-\alpha)$  quantile of the distribution of  $\hat{w}_n(\delta_0)$  under the null, evaluated by Monte Carlo simulations, at some  $\delta_0$  specified *a priori*.

## 5. Monte Carlo experiments.

5.1. *Experimental setup.* In this Section we compare numerically the tests defined in Section 4 that are denoted MULTIPLE, PLUGIN, NN, and TwoPC.

For  $T$  being any of those non randomized test procedures, we can tune the parameters of the procedure to have a prescribed level  $\alpha$ , *i.e.*  $\mathbb{P}_g(T = 1) = \alpha$ .

This is done by Monte Carlo replication. Ten thousand independent random samples of size  $n$  are drawn under the null hypothesis, for  $g$  being the uniform density on  $\mathbb{S}$  ( *i.e.*  $g \equiv 1/(4\pi)$ ) or the stylized coverage function of the Pierre Auger detector (see Figure 1).

For the MULTIPLE procedure and a given level  $\alpha$ , we have chosen

$$T_j = \mathbf{1}_{\|\hat{f}_j - g\|_p > t_{\alpha', j}} \quad (22)$$

where  $t_{\alpha', j}$  is the  $1 - \alpha'$  quantile of the distribution of  $\|\hat{f}_j - g\|_p$  under the null hypothesis. This distribution is evaluated using Monte Carlo replications. Further, the value  $\alpha'$  is chosen so that

$$T'_{J^*} = \sup_{j=1, \dots, J^*} T_j \quad (23)$$

has a first type error probability equal to  $\alpha$ . This is arbitrary and the theory to be written would likely suggest to use a scale dependent level.

The power of the test  $T$  is defined by (7). Some clues about this value are obtained by evaluating  $\mathbb{P}_f(T = 1)$  for particular alternatives  $f$  that are given in the next Section. Here again, those quantities are evaluated by Monte Carlo. Note however that the power for a particular alternative only gives an upper bound of the power in the minimax sense given by the second equation of (7).

In the following tables of tests we represent the power of the needlet tests as a function of the finest band  $J^*$  and the power of the norm we use to detect anisotropy (see Appendix A for more details on the actual implementation of the method).

The profile cuts of the (axisymmetric) needlets we have used are plotted the online supplement.

**5.2. Alternatives.** We have investigated the performance of the test (power against level) for sample sets of small to moderate size ( $n = 25, 100, 400$ ) and against different alternatives. Those choices of  $n$  mimics the progressive publication of events by the Pierre Auger Observatory (27 events above  $5.7 \times 10^{19}$  eV in 2008, 69 above  $5.5 \times 10^{19}$  in 2010, a few hundreds in the future).

Generally speaking, the physical plausibility of those alternatives is weak (alternative  $(H_1^c)$ ), if not null (alternatives  $(H_1^b)$  and  $(H_1^c)$ ). Our goal is to focus here on specific departures from isotropy. First we consider unimodal non isotropic densities, with a Gaussian shape. Then we consider mixtures of densities that would only be obtained if the sources of the cosmic rays were known to be uniformly distributed and repeating, and at the same

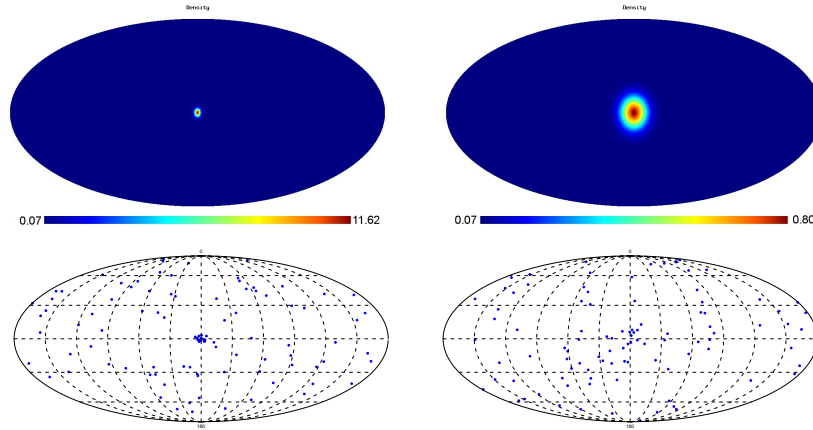


Fig 3: Densities (first line) and random draws (second line,  $n = 100$ ) under  $(H_1^a)$  with  $\delta = 10\%$  and  $\theta = 5^\circ$  (left) or  $\theta = 20^\circ$  (right).

distance from us. Third, the Physics-inspired model  $(H_1^c)$  give rise to non-isotropic patterns with richer frequency content compared to the previous ones (and non axisymmetric clusters). We now give the precise definitions of the alternatives.

$(H_1^a)$ . The first family of alternatives is obtained as a mixture of the uniform density  $f_0$  and an over-density at some point of the sphere, with Gaussian-like axisymmetric profile. Precisely, the density under  $(H_1^a)$  writes

$$f(\xi) = (1 - \delta)f_0 + \delta h_\theta(\xi)$$

where  $h_\theta(\xi) := h_{\theta, \xi_0}(\xi)$ ,  $h_{\theta, \xi_0} := C_\theta \exp(-(\xi \cdot \xi_0)^2 / 2\theta^2)$  and  $\xi_0 = (\pi/2, 0)$ . Such densities are then unimodal, with a bump whose width is proportional to  $\theta$ . Typical observations of random draw with such density with  $\delta = 0.01$  and  $\theta = 5^\circ$  or  $20^\circ$  are displayed on Figure 3.

$(H_1^b)$ . A second family of alternative is a toy model for the repeating emission of events from a small number of sources, as explained in the Introduction. Here we assume that the  $n_s$  sources are uniformly distributed, although in a realistic case, we can expect any type of astrophysical sources to follow the local matter density of the cosmic structure (which would make the detection of anisotropy easier). This generalisation is straightforward enough that we do not discuss it further at this stage. Conditionally to those positions, the  $X_i$  are distributed along a mixture of  $n_s$  Gaussian densities centred on the sources (to take into account the error in the measurement of the incidence angle or the deflection of the charged particle by Galactic magnetic



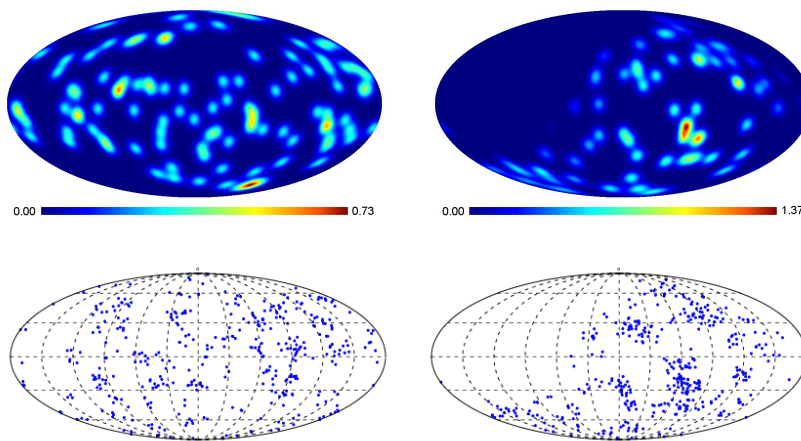


Fig 4: Density of  $X_1$  conditionally to the random draw of the centers of 100 AGNs for a uniform coverage (first line) and random draws with  $n = 400$  (second line). The coverage is uniform on the left, *à la* South Pierre-Auger observatory on the right.

fields) multiplied by the coverage function ( $g$ ) of the instrument. Namely

$$f(\xi) = g(\xi) \sum_{j=1}^{n_s} h_{\theta, \xi_i}(\xi) .$$

Such conditional densities are displayed on the first line of Figure 4. We considered the cases  $n_s = 10$  and  $n_s = 100$  and fixed  $\theta = 10^\circ$ . Note that if  $n_s$  is much bigger than  $n$ , it is difficult to detect this kind of anisotropy (which can be detected only if at least one source has emitted more than one cosmic ray).

( $H_1^c$ ). A third and last alternative is obtained by the physical model of cosmic ray observations described in detail in Section 2. Sources are randomly drawn in a spherical volume of radius  $r_{\max} = 70$  Mpc, and their flux is assumed inversely proportional to the square of their distance. The parameters for the simulations are taken to be  $E_{\max} = 10^{21}$  eV,  $\alpha = 4.2$ . We consider different values for  $E_{\min}$  (namely 1, 4 or  $6 \times 10^{19}$  eV). Playing on this parameter has an important practical incidence. Assuming that the distribution of the energy of the cosmic rays is a power law,  $\mathbb{P}(E > t) \sim Ct^{-\alpha+1}$ , lowering the threshold on the selection of the cosmic rays from  $6 \times 10^{19}$  eV to  $4 \times 10^{19}$  eV (resp  $10^{19}$  eV) accounts to increase the size of the sample (available observations above the threshold) by a factor  $(6/4)^{\alpha-1} \simeq 3.66$  (resp 310). It means

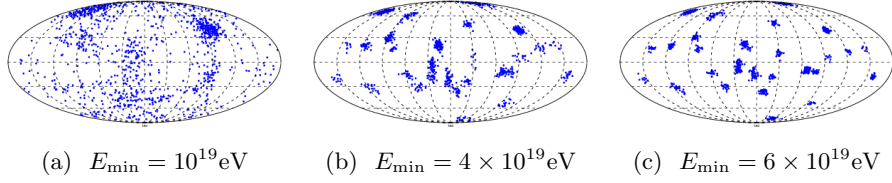


Fig 5: Isotropization of the cosmic rays in model  $(H_1^c)$  as  $E_{\min}$  decreases. There are the exact same  $n_s = 30$  sources in the three cases and  $n = 1000$  observations.

that the statistical decision should be made far easier if the cosmic rays were not too much isotropized by the Galactic fields as their energies goes lower. This effect is illustrated in Figure 5. It is interesting to see if the methods are still able to detect anisotropy as the cosmic rays become more and more isotropized. This is a more realistic simulation compared to models  $(H_1^a)$  and  $(H_1^b)$ . There is no single size for the scatter of the CRs coming for a given source, nor the same size or directionality for each source, nor the same flux for each source, that hence is interesting specifically for a multiscale analysis with no prior assumption about a correlation length.

Note that under the alternatives  $(H_1^b)$  and  $(H_1^c)$ , the procedure is to be understood as a test on the conditional distribution of  $(X_i)_{i=1,\dots,n}$  with respect to the positions of the “sources”, which are randomly drawn once for all.

### 5.3. Numerical results and discussion.

*Tables.* We shall represent some of the results of our simulations with tables of estimated power of the procedures for given alternatives (in percent), at the prescribed level  $\alpha = 0.05$ . Practically, we let the finest needlet band entering the MULTIPLE and PLUGIN procedures vary in the set  $\{J^* - 2, J^* - 1, J^*, J^* + 1\}$  where  $J^*$  is given by (14). The entry (or entries) corresponding to the overall highest power (before rounding off) among the 26 values is (are) printed in bold type. We consider three  $L^p$  norms, namely  $L^p$  for  $p = 1, 2, \infty$ . It is possible to use an unbiased estimate of the distance between  $\hat{f}$  and  $g$  in the case of the  $L^2$  norm. It is referred to as  $p = 2^*$  (see Appendix A.2.3 in the on-line supplement for details)

*ROC curves.* The receiver operating characteristic (ROC) curves plot the power  $p$  of a procedure as a function of its level  $\alpha$ . It is a useful representation for comparison of different procedures along a wide range of levels. The ROC

curves associated to the TwoPC procedure are step function because of the discrete nature of the test statistic. Some of the ROC curves are non concave. It should be recalled however that any procedure of this kind can be improved to a randomized procedure whose ROC curve is the concave upper envelope of the original one. Accordingly, the reader's eyes must actually analyse the upper envelopes of the ROC curves. Note that the power in the tables have not been modified by this argument.

ROC curves are represented in plots with four subplot, corresponding to the four above-mentioned choices of  $J^*$  in the needlet methods. The ROC curves for TwoPC and NN procedures are the same in the four subplots. Inset graphs allow complementary comparison of the methods by zooming on the most relevant levels (small  $\alpha$ ).

**5.3.1. Some specific results.** Figure 7 illustrates that the methods MULTIPLE and PLUGIN have a consistent behaviour when the typical radius of anisotropic structure is varying. The model we consider here is  $(H_1^a)$  with  $\theta = 5^\circ$  or  $\theta = 20^\circ$ . We shall discuss further from those cases below. Figure 8 illustrate the good performances even for small samples under the model  $(H_1^c)$  that produces clusters of various sizes and shapes.

The differences of sensitivity between the different norms are not very strong, probably because we consider quite regular alternative hypotheses. As expected, the  $L^\infty$  is a bit more sensitive to more spiky (unimodal) distributions, whereas more global measures such as  $L^1$  or  $L^2$  perform better under the  $(H_1^b)$  or  $(H_1^c)$  models. This is illustrated by the ROC curves of Figure 13 in the online supplement.

Under the alternative  $(H_1^b)$  the NN procedure performs strikingly well, as illustrated by the power of the four procedures in Table 5. Recall that there is no parameter to tune in this method. Those good results can be explained in the following manner. Under  $(H_1^b)$ , the points  $\{X_i\}_{i=1,\dots,n}$  are mainly grouped into clusters of average scale given by the standard deviation of the Gaussians of the mixture. If the number of clusters and this standard deviation are too small to cover significantly the whole sphere, then the random distances to the nearest neighbour are bounded by  $\sigma$  with very high probability, which is not the case under the null. This makes the distribution of the distance to the nearest neighbour a very sensitive tool to discriminate between  $(H_1^b)$  and the null.

Varying the alternatives, it appears that no method outperforms the other in a uniform way, but it seems that the two needlet methods, if not always optimal, consistently have a good behaviour. Moreover, the MULTIPLE test is slightly more sensitive than the PLUGIN one. As an illustration, we represent

in Tables 3 to 8 the results of the simulation for a quite representative panel of alternatives, containing more or less spiky distributions, clusters of smoother alternatives, weak or strong anisotropy etc.

**5.3.2. Sample size.** It is important to notice that the TwoPC approach often provides a good sensitivity if not the best at  $n = 25$ . For most of the alternative, however, one or the other of the needlets methods outperforms TwoPC as  $n$  grows. This is exemplified in Tables 3 to 8. For  $n = 100$ , the needlet methods are more sensitive, whereas for  $n = 25$  they are usually less sensitive than TwoPC. As already noticed, NN is consistently better for the model  $(H_1^b)$  (see Table 5).

In our application context, the sample size over a given energy threshold is increasing with time and experiments, so it must be highlighted that multiscale methods are more and more appropriate for analysis of future datasets.

**5.3.3. Separation rate.** We focus here on the behavior of the power of the test with respect to  $n$ . If  $r_n$  is the critical rate in the minimax sense (given by Equations 8 and 9), we should observe an approximately same power for different sample size and the least favorable alternative densities  $\tilde{f}_n$  as soon as the quantity  $r_n d(\tilde{f}_n, g)$  remains constant. On Table 9, we have displayed the power of the different procedures for three different densities corresponding to the alternative  $(H_1^a)$  and three sample sizes, keepin the same value for  $n^{1/2} d(f, g)$ . Indeed, in the  $(H_1^a)$  case, for any power norm,  $d(f, f_0) = \delta d(f_0, h_\theta)$ . As the power remains roughly in  $(0, 1)$  the same for the three values of the parameters, and as  $n^{1/2}$  is an upper bound four the minimax separation rate in analogy with similar problems on Euclidean spaces, this numerical simulations is consistent with the claim that the needlet based procedures perform well at the minimax rate of testing. The increasing value of the power with  $n$  together with the unbeatable rate of separation  $\sqrt{n}$  illustrates the fact that we only have access to upper bounds of the minimax rate. In other words, the densities under consideration are definitely not the least favorable cases. The comparison of needlet methods with NN and TwoPC methods tends to be more favourable to needlets methods as  $n$  becomes larger in this case.

**5.3.4. Robustness.** Assume that the anisotropy detection by the needlet methods is adaptive. Then, as pre-tuned black boxes, those methods should remain optimal on a wide range of alternatives. Some simulations support this claim. Note however that we only explore physically possible alternative which are smooth non uniform densities.

The key parameter of the TwoPC method is the angular size  $\delta_0$  at which we compare  $\hat{w}(\delta)$  to the distribution of  $w(\delta_0)$  under the null. For sake of fairness in our comparisons, we should allow some tuning of this parameter. It is clear that the optimal  $\delta_0$  is related to the “average scale” of the anisotropy. Though it is difficult to give a precise and general definition of this former quantity, it should be close to the value of the parameter  $\theta$  in the particular case of model  $(H_1^a)$ . Figure 7 has been obtained for two densities of this model, with  $\delta_0 = 3^\circ$ . It appears that TwoPC is better than the needlet methods when  $\theta = 5^\circ$  and worse when  $\theta = 20^\circ$ .

On Figure 9 we have plotted the estimated power of the tests against different alternative  $(H_1^a)$ , with different parameters, and for different parameters for the methods. The figure shows that the optimal  $\delta_0$  is indeed related to the parameters  $\theta$  of the alternative. However, when dealing with alternative that give rise to structures at different “scales”, the optimal choice of  $\delta_0$  is not clear. This is illustrated by Figure 10, where models  $(H_1^c)$  are under consideration. By observing the large variations of the power of the TwoPC procedure with respect to  $\delta_0$  in both cases, one can conclude that this procedure should incorporate a data-driven selection of  $\delta_0$  to be truly efficient.

The situation is strikingly different for the needlet methods. One can observe from the left columns of Figures 9 and 10 that the power reaches some *plateau* after  $J^* > J_{\min}$  in a very consistent way across the different alternatives. This robustness is a strong point of those methods. The dependence in  $n$  is quite weak too. For instance, taking  $J^* = 4$  leads to a small loss of efficiency uniformly with respect to the best choice for each given situation of sample size and model.

**6. Analysis of Auger data.** We have run the previous tests on the Auger public data made available by (The Pierre AUGER Collaboration et al., 2010). It is composed of 69 arrival directions of cosmic rays with energy above 55 EeV and detected by the Pierre Auger Observatory between 1 January 2004 and 31 December 2009. Those directional events are plotted on Figure 6. The distributions of the tests under study for  $n = 69$  and under the null hypothesis has been evaluated by Monte-Carlo simulation of length 10.000.

Along with the detection of a correlation between cosmic rays directions and catalogues of potential sources, the Pierre Auger collaboration already performed a catalogue-free test for anisotropy with no reference to any catalogue, using the TwoPC procedure. As noticed earlier, the critical value for this method is the choice of  $\delta_0$  in (21). The  $p$ -value of this test for the

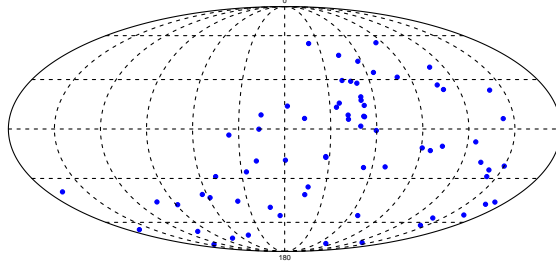


Fig 6: The 69 arrival directions of cosmic rays with energy above 55 EeV and detected by the Pierre Auger Observatory up to 31 December 2009 ([The Pierre AUGER Collaboration et al., 2010](#))

69 UHECRs data set reaches a minimal value of

$$p\text{-value}(\text{TwoPC}) \simeq 0.008$$

around  $\delta_0 \simeq 10.7^\circ$ . Recall that in order to be interpretable as a classical  $p$ -value for a single hypothesis testing, this  $p$ -value should have been computed from an out-of-the-sample prescription of  $\delta_0$ , which is not the case here. Then this  $p$ -value strongly exaggerates the significance of the detection. Indeed, as already noticed in ([The Pierre AUGER Collaboration et al., 2010](#)), we computed that the fraction of isotropic simulations that are as non-isotropic as the real data at *some* angle between  $4^\circ$  and  $14^\circ$  is as high as 10%. We have also computed that

$$p\text{-value}(\text{NN}) \simeq 0.07 .$$

The  $p$ -values of Table 1 are the  $p$ -values computed from the Pierre Auger data set for our MULTIPLE and PLUGIN procedure.

For the MULTIPLE test, the  $p$ -value is defined as the proportion of draws (under the null) that have a higher single test statistic in at least one value of  $j \in \{1, \dots, J^*\}$ . The resulting  $p$ -value is quite sensitive to the choice of the highest band  $J^*$ , except if one uses the  $L^2$ -norm. Note that if we take the  $L^2$  norm and the theoretical  $J^* = 2$  given by the expression (14), the results for the MULTIPLE test are not statistically significant. But the Monte-Carlo simulations suggest that this theoretical choice of  $J^*$  is not optimal for small to medium sample size, being too small.

The PLUGIN is more stable and consistently considers that the Auger data is significantly non isotropic. The almost constant  $p$ -values in this case

$J^*$	$p = 1$	$p = 2^*$	$p = \infty$	$J^*$	$p = 1$	$p = 2^*$	$p = \infty$
1	0.957	0.788	0.387	1	0.956	0.958	0.397
2	0.051	0.112	0.035	2	0.033	0.037	0.036
3	0.118	0.050	0.004	3	0.017	0.008	0.005
4	0.434	0.046	0.003	4	0.017	0.008	0.008
5	0.227	0.095	0.624	5	0.017	0.008	0.008
6	0.762	0.045	0.341	6	0.017	0.008	0.008

(a) MULTIPLE test

(b) PLUGIN test

TABLE 1

*P-values of the MULTIPLE and the PLUGIN tests for Auger data ( $n = 69$ ).*

is the consequence of a hard thresholding rule in (15) that cancels all the estimated coefficients  $\hat{\beta}_{j,k}$  as soon as  $j \geq 3$  for this data set. This may in turn give a rule-of-thumb rule to define a data-driven  $J^*$  for the multiple test.

To conclude on this important data set and this methodology, it appears that the needlet methods find a stronger statistical evidence of some kind of anisotropy in the Pierre Auger data. More realistic alternatives and more simulations can help to choose the  $J^*$  parameter of the MULTIPLE procedure, and additional parameters of the PLUGIN approach.

**7. Conclusion.** In this paper, we have investigated the problem of the detection of anisotropy of directional data on the unit sphere, with an application to the analysis of ultra high energy cosmic ray events as observed with a detector such as the Pierre Auger observatory. It was important to consider samples whose sizes are comparable to the sizes of the data sets that are available nowadays for cosmic rays scientists (about 25 at the beginning of this work, about a hundred now). Although we are mainly interested in small sample performances, we have proposed a multiple test approach based on a multiresolution analysis of the data, which could hopefully be proved to be asymptotically optimal in the minimax sense, a well-known pessimistic framework.

We have proposed, and tested on various simulated data sets, two methods using the decomposition of the directional data onto a frame of spherical needlets. Their performance has been compared to other (more specific) approaches based on the nearest neighbour and on the two point correlation function. The simulation shows that the needlet-based methods perform comparatively very well in various situations. They are competitive with existing method at small sample size, and tend to outperforms them from moderate sample size. Moreover, the “omnibus” property of the needlets method is interesting for the problem at hand, in which the type of pos-



sible anisotropy (the class of alternative) is not really well known a priori. In addition, a multiple test based on the use of spherical needlets offers a good opportunity to extend the method of detection of anisotropies with not only multiplicity in the scales tested, but also in ranges of energy of the incoming particles. Indeed, while in this work we have used the energy level as a simple threshold, one could instead implement a detection using the joint directional-energy information – allowing thus to simultaneously extract information from the highest energy cosmic rays, which are not deflected much by Galactic and extragalactic magnetic fields, and also from lower energy events, more deflected but much more numerous. In the light of our simulations on an energy level-dependent model, multiscale approach could lead to stronger conclusion using the CR data that are not yet made public by the Pierre Auger Observatory.

As in any nonparametric method, there is at least one parameter to be tuned, often by hand or using more sophisticated data-driven methods such as cross-validation. In the needlet methods one can tune several parameters (shape of the needlets, highest scale  $J^*$ — although there’s is an asymptotic formula for it—, thresholds on the coefficients in the `PLUGIN` approaches, thresholds on the individual tests in the `MULTIPLE` procedure, power norm). It is plausible however that a large range of possible choices for most of these parameters give comparable performance.

Although we have used needlets that are compactly supported windows in the harmonic space, it may be arguable that they are not the most appropriate tool. One could consider, as an alternative, better spatially concentrated functions (such as the Mexican needlets, see *e.g.* [Lan and Marinucci, 2009](#)) or, in general, try to optimize the needlet window function given prior knowledge of the physical problem and of the expected properties of anisotropic distributions of the cosmic ray direction of incidence. In this spirit, it would be interesting to consider directional wavelet such as curvelets or ridgelets (see [Starck et al., 2006](#)) to test for specific strip-like alternative densities. It is also possible to consider non-dyadic needlets. The choice of  $\mathcal{B} \in (1, 2)$  allows a finer coverage of the frequency line. The numerical results presented here have not taken this benefit into full account, and whether significantly higher power can be obtained by optimizing this number remains to be investigated.

Finally, in addition to the aforementioned possible extensions of our methods, we want to stress that the work presented here also opens the way to two lines of future investigations, one on the applications side, and one more theoretical. On the experimental side, it will be of much interest to apply the method on larger data sets (for instance by lowering the energy threshold

to increase the available sample size). On the theoretical side, the validation of the approach has to be investigated on the basis of some theory in the minimax framework it is designed for.

*Acknowledgment.* For the numerical part of this work, we acknowledge the use of the HEALPix package (Górski et al., 2005) and of the SphereLab (an OCTAVE interface to HEALPix package, needlet transforms, and utilities). We thank Dmitri Semikoz for useful discussions concerning high energy cosmic rays, and Maude le Jeune, Jean-François Cardoso and Frédéric Guillaux for making available some of the tools used for the computational aspects of the present work. We thank the anonymous referees and associate editor for their suggestions and remarks that improved the presentation of this paper.

## References.

- J. Abraham et al. Observation of the Suppression of the Flux of Cosmic Rays above  $4 \times 10^{19}$  eV. *Physical Review Letters*, 101(6):061101–+, August 2008. .
- P. Baldi, G. Kerkycharian, D. Marinucci, and D. Picard. Asymptotics for spherical needlets. On arXiv.org : math.ST/0606599, 2006.
- P. Baldi, G. Kerkycharian, D. Marinucci, and D. Picard. Subsampling Needlet Coefficients on the Sphere. *ArXiv e-prints*, 706, June 2007.
- P. Baldi, G. Kerkycharian, D. Marinucci, and D. Picard. Asymptotics for spherical needlets. *Ann. Statist.*, 2008. (In press).
- P. Baldi, G. Kerkycharian, D. Marinucci, and D. Picard. Adaptive density estimation for directional data using needlets. *Ann. Statist.*, 37(6A):3362–3395, 2009. ISSN 0090-5364.
- Pijushpani Bhattacharjee and Guenter Sigl. Origin and propagation of extremely high energy cosmic rays. *Physics Reports*, 327:109–247, March 2000. .
- Cristina Butucea and Karine Tribouley. Nonparametric homogeneity tests. *J. Statist. Plann. Inference*, 136(3):597–639, 2006. ISSN 0378-3758.
- Ingrid Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
- J. Delabrouille, J.-F. Cardoso, M. Le Jeune, M. Betoule, G. Faÿ, and F. Guillaux. A full sky, low foreground, high resolution CMB map from WMAP. *Astron. Astrophys.*, 493: 835–857, January 2009. .
- David L. Donoho, Iain M. Johnstone, Gérard Kerkycharian, and Dominique Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2):508–539, 1996. ISSN 0090-5364. . URL <http://dx.doi.org/10.1214/aos/1032894451>.
- B. Efron and V. Petrosian. Testing Isotropy versus Clustering of Gamma-Ray Bursts. *Astrophys. J.*, 449:216–+, August 1995. .
- G. Faÿ and F. Guillaux. Consistency of a needlet spectral estimator on the sphere. *Statistical Inference for Stochastic Processes*, 14:47–71, 2011.
- G. Faÿ, F. Guillaux, M. Betoule, J.-F. Cardoso, J. Delabrouille, and M. Le Jeune. CMB power spectrum estimation using wavelets. *Physical Review D (Particles, Fields, Gravitation, and Cosmology)*, 78(8):083013, 2008. .
- Magalie Fromont and Béatrice Laurent. Adaptive goodness-of-fit tests in a density model. *Ann. Statist.*, 34(2):680–720, 2006. ISSN 0090-5364.

- Evarist Giné. Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.*, 3(6):1243–1266, 1975. ISSN 0090-5364.
- K. Górski, E. Hivon, A. Banday, B. Wandelt, F. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *Astrophys. J.*, 622:759–771, April 2005. . Package available at <http://healpix.jpl.nasa.gov>.
- K. Greisen. End to the Cosmic-Ray Spectrum? *Physical Review Letters*, 16:748–750, April 1966. .
- F. Guilloux, G. Faÿ, and J.-F. Cardoso. Practical wavelet design on the sphere. *Appl. Comput. Harmon. Anal.*, 26(2):143–160, march 2009.
- D. Harari, S. Mollerach, and E. Roulet. Astrophysical magnetic field reconstruction and spectroscopy with ultra high energy cosmic rays. *Journal of High Energy Physics*, 7: 6–+, July 2002. .
- A. M. Hillas. The Origin of Ultra-High-Energy Cosmic Rays. *Annual Review of Astronomy and Astrophysics*, 22:425–444, 1984. .
- Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. Methods Statist.*, 2(2):85–114, 171–189, 249–268, 1993.
- Yu. I. Ingster. Adaptive chi-square tests. *J. Math. Sciences*, 99:1110–1120, 2000.
- Yu. I. Ingster and I. A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2003. ISBN 0-387-95531-3.
- Anatoli Juditsky and Sophie Lambert-Lacroix. On minimax density estimation on  $\mathbb{R}$ . *Bernoulli*, 10(2):187–220, 2004. ISSN 1350-7265.
- Michael Kachelriess and D. V. Semikoz. Clustering of ultra-high energy cosmic ray arrival directions on medium scales. *Astropart. Phys.*, 26:10–15, 2006. .
- G Kerkycharian and D. Picard. New generation wavelets associated with statistical problems. In *The 8th Workshop on Stochastic Numerics*, pages 119–146. Research Institute for Mathematical Sciences, Kyoto University, Jan. 2009.
- G rard Kerkycharian, Pencho Petrushev, Dominique Picard, and Thomas Willer. Needlet algorithms for estimation in inverse problems. *Electron. J. Stat.*, 1:30–76 (electronic), 2007. ISSN 1935-7524.
- G rard Kerkycharian, George Kyriazis, Erwan Le Pennec, Pencho Petrushev, and Dominique Picard. Inversion of noisy Radon transform by SVD based needlets. *Appl. Comput. Harmon. Anal.*, 28(1):24–45, 2010. ISSN 1063-5203. .
- Gerard Kerkycharian, Thanh Mai Pham Ngoc, and Dominique Picard. Localized deconvolution on the sphere. *The Annals of Statistics*, 39(2):1042–1068, 2011.
- Xiaohong Lan and Domenico Marinucci. On the dependence structure of wavelet coefficients for spherical random fields. *Stochastic Process. Appl.*, 119(10):3749–3766, 2009. ISSN 0304-4149. . URL <http://dx.doi.org/10.1016/j.spa.2009.07.005>.
- E. L. Lehmann and Joseph P. Romano. *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005. ISBN 0-387-98864-5.
- D. Marinucci, D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkycharian, P. Natoli, D. Picard, and N. Vittorio. Spherical Needlets for CMB Data Analysis. *M.N.R.A.S.*, 383(2):539–545, January 2008.
- V. J. Mart nez and E. Saar. *Statistics of the Galaxy Distribution*. Chapman & Hall/CRC, 2002.
- Yves Meyer. *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1992. ISBN 0-521-42000-8; 0-521-45869-2. Translated from the 1990 French original by D. H. Salinger.
- F. Narcowich, P. Petrushev, and J. Ward. Decomposition of Besov and Triebel-Lizorkin

- spaces on the sphere. *J. Funct. Anal.*, 238:530–564, September 2006a.
- F. J. Narcowich, P. Petrushev, and J. D. Ward. Localized tight frames on spheres. *SIAM J. Math. Anal.*, 38(2):574–594, 2006b.
- P. Petrushev and Y. Xu. Localized polynomials frames on the ball. *Constr. Approx.*, 27:121–148, 2008.
- D. Pietrobon, A. Amblard, A. Balbi, P. Cabella, A. Cooray, and D. Marinucci. Needlet detection of features in the WMAP CMB sky and the impact on anisotropies and hemispherical asymmetries. *Physical Review D*, 78(10):103504–+, November 2008. .
- Davide Pietrobon, Amedeo Balbi, and Domenico Marinucci. Integrated Sachs-Wolfe effect from the cross correlation of WMAP3 year and the NRAO VLA sky survey data: New results and constraints on dark energy. *Phys.Rev. D*, 74, 2006.
- J. M. Quashnock and D. Q. Lamb. Evidence for the Galactic origin of gamma-ray bursts. *M.N.R.A.S.*, 265:L45–L50, December 1993.
- Ø. Rudjord, F. K. Hansen, X. Lan, M. Liguori, D. Marinucci, and S. Matarrese. An Estimate of the Primordial Non-Gaussianity Parameter  $f_{\text{NL}}$  Using the Needlet Bispectrum from WMAP. *The Astrophysical Journal*, 701:369–376, August 2009. .
- V. G. Spokoiny. Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24(6):2477–2498, 1996. ISSN 0090-5364. . URL <http://dx.doi.org/10.1214/aos/1032181163>.
- J.-L. Starck, Y. Moudden, P. Abrial, and M. Nguyen. Wavelets, ridgelets and curvelets on the sphere. *Astronomy & Astrophysics*, 446:1191–1204, 2006.
- The Pierre AUGER Collaboration, J. Abraham, et al. Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei. *Astroparticle Physics*, 29:188–204, April 2008. .
- The Pierre AUGER Collaboration, P. Abreu, M. Aglietta, E. J. Ahn, D. Allard, I. Allekotte, J. Allen, J. Alvarez Castillo, J. Alvarez-Muñiz, M. Ambrosio, and et al. Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter. *Astroparticle Physics*, 34:314–326, December 2010. .
- D. F. Torres and L. A. Anchordoqui. Astrophysical origins of ultrahigh energy cosmic rays. *Reports on Progress in Physics*, 67:1663–1730, September 2004. .
- Hans Triebel. *Theory of function spaces. II*, volume 84 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1992. ISBN 3-7643-2639-5. .
- G. T. Zatsepin and V. A. Kuz'min. Upper Limit of the Spectrum of Cosmic Rays. *Soviet Journal of Experimental and Theoretical Physics Letters*, 4:78–+, August 1966.

$J^*$		3	4	5	6	$J^*$		3	4	5	6
MULTIPLE	$p = 1$	77	<b>100</b>	<b>100</b>	<b>100</b>	MULTIPLE	$p = 1$	96	<b>100</b>	<b>100</b>	<b>100</b>
	$p = 2^*$	71	<b>100</b>	<b>100</b>	<b>100</b>		$p = 2^*$	78	100	<b>100</b>	100
	$p = \infty$	26	36	42	43		$p = \infty$	56	66	66	63
PLUGIN	$p = 1$	61	61	61	61	PLUGIN	$p = 1$	68	68	68	68
	$p = 2$	60	60	60	60		$p = 2$	68	72	72	72
	$p = \infty$	31	36	36	36		$p = \infty$	42	52	52	52
NN		99				NN		100			
TwoPC		60				TwoPC		80			

(a) Uniform exposure
(b) Auger exposure

TABLE 2

Power (in %) under  $(H_1^b)$  with number of sources  $n_s = 100$ , sample size  $n = 25$ , and for tests of size  $\alpha = 0.01$ . The increase of power can be understood by the growing effective density of observations, at fixed sample size, when restricting the field of observation from the whole sphere to the Pierre-Auger acceptance

$J^*$		3	4	5	6	$J^*$		3	4	5	6
MULTIPLE	$p = 1$	10	14	12	10	MULTIPLE	$p = 1$	20	30	33	28
	$p = 2^*$	10	12	18	18		$p = 2^*$	27	51	66	69
	$p = \infty$	10	15	20	20		$p = \infty$	40	68	76	<b>77</b>
PLUGIN	$p = 1$	10	10	10	10	PLUGIN	$p = 1$	22	20	19	19
	$p = 2$	12	13	13	13		$p = 2$	36	48	50	50
	$p = \infty$	16	20	20	20		$p = \infty$	38	71	71	71
NN		9				NN		12			
TwoPC		<b>46</b>				TwoPC		76			

(a)  $n = 25$ 
(b)  $n = 100$

TABLE 3

Power (in %) under  $(H_1^a)$ , under Pierre-Auger exposure, with  $\delta = 0.04$  and  $\theta = 5$ .

$J^*$		3	4	5	6	$J^*$		3	4	5	6
MULTIPLE	$p = 1$	14	18	21	20	MULTIPLE	$p = 1$	40	44	59	54
	$p = 2^*$	17	22	25	26		$p = 2^*$	61	77	86	88
	$p = \infty$	15	24	30	30		$p = \infty$	69	89	<b>94</b>	<b>94</b>
PLUGIN	$p = 1$	10	10	10	10	PLUGIN	$p = 1$	45	43	43	43
	$p = 2$	12	16	17	17		$p = 2$	64	74	76	76
	$p = \infty$	22	22	22	22		$p = \infty$	74	88	88	88
NN		9				NN		19			
TwoPC		<b>50</b>				TwoPC		84			

(a)  $n = 25$ 
(b)  $n = 100$

TABLE 4

Power (in %) under  $(H_1^a)$ , under Pierre-Auger exposure, with  $\delta = 0.08$  and  $\theta = 5$ .

$J^\star$		3	4	5	6
MULTIPLE	$p = 1$	32	42	38	35
	$p = 2^\star$	26	38	39	38
	$p = \infty$	19	26	29	30
PLUGIN	$p = 1$	7	6	6	6
	$p = 2$	14	15	15	15
	$p = \infty$	26	26	26	26
NN		28			
TwoPC		<b>60</b>			

(a)  $n = 25$

$J^\star$		3	4	5	6
MULTIPLE	$p = 1$	100	<b>100</b>	<b>100</b>	<b>100</b>
	$p = 2^\star$	96	<b>100</b>	<b>100</b>	<b>100</b>
	$p = \infty$	80	85	86	84
PLUGIN	$p = 1$	90	91	91	91
	$p = 2$	88	91	91	91
	$p = \infty$	62	78	78	78
NN		<b>100</b>			
TwoPC		92			

(b)  $n = 100$

TABLE 5

Power (in %) under  $(H_1^b)$ , under Pierre-Auger exposure, with  $M = 100$  and  $\theta = 10$ .

$J^\star$		3	4	5	6
MULTIPLE	$p = 1$	51	46	41	40
	$p = 2^\star$	52	<b>53</b>	47	47
	$p = \infty$	42	44	42	42
PLUGIN	$p = 1$	34	34	34	34
	$p = 2$	42	42	42	42
	$p = \infty$	50	50	50	50
NN		38			
TwoPC		45			

(a)  $n = 25$

$J^\star$		3	4	5	6
MULTIPLE	$p = 1$	98	98	98	98
	$p = 2^\star$	98	<b>99</b>	98	98
	$p = \infty$	92	91	91	90
PLUGIN	$p = 1$	98	98	98	98
	$p = 2$	98	98	98	98
	$p = \infty$	92	92	92	92
NN		82			
TwoPC		62			

(b)  $n = 100$

TABLE 6

Power (in %) under  $(H_1^c)$ , under uniform exposure, with  $n_s = 100$  and  $E_{\min} = 10^{19}$ .

$J^\star$		3	4	5	6
MULTIPLE	$p = 1$	30	40	41	40
	$p = 2^\star$	34	46	46	47
	$p = \infty$	29	36	44	47
PLUGIN	$p = 1$	18	18	18	18
	$p = 2$	26	26	26	26
	$p = \infty$	33	35	35	35
NN		34			
TwoPC		<b>71</b>			

(a)  $n = 25$

$J^\star$		3	4	5	6
MULTIPLE	$p = 1$	85	94	98	98
	$p = 2^\star$	87	98	<b>99</b>	<b>99</b>
	$p = \infty$	74	90	95	96
PLUGIN	$p = 1$	72	72	72	72
	$p = 2$	75	77	78	78
	$p = \infty$	68	76	76	76
NN		95			
TwoPC		96			

(b)  $n = 100$

TABLE 7

Power (in %) under  $(H_1^c)$ , under uniform exposure, with  $n_s = 500$  and  $E_{\min} = 10^{19}$ .

	$J^*$	3	4	5	6
MULTIPLE	$p = 1$	27	39	45	43
	$p = 2^*$	28	42	50	58
	$p = \infty$	24	35	45	50
PLUGIN	$p = 1$	18	18	18	18
	$p = 2$	25	28	29	29
	$p = \infty$	33	34	34	34
NN		33			
TwoPC		<b>75</b>			

(a)  $n = 25$ 

	$J^*$	3	4	5	6
MULTIPLE	$p = 1$	76	94	99	98
	$p = 2^*$	79	96	100	<b>100</b>
	$p = \infty$	69	84	96	97
PLUGIN	$p = 1$	72	72	73	73
	$p = 2$	78	82	82	82
	$p = \infty$	71	80	80	80
NN		99			
TwoPC		99			

(b)  $n = 100$ 

TABLE 8

Power (in %) under  $(H_1^c)$ , under uniform exposure, with  $n_s = 500$  and  $E_{\min} = 5.10^{19}$ .

	$J^*$	3	4	5	6
MULTIPLE	$p = 1$	14	16	13	13
	$p = 2^*$	19	20	16	16
	$p = \infty$	23	26	23	22
PLUGIN	$p = 1$	11	11	11	11
	$p = 2$	16	16	16	16
	$p = \infty$	23	22	22	22
NN		8			
TwoPC		<b>35</b>			

(a)  $n = 25, \delta = 0.08$ 

	$J^*$	3	4	5	6
MULTIPLE	$p = 1$	14	16	14	14
	$p = 2^*$	17	21	20	20
	$p = \infty$	29	32	32	30
PLUGIN	$p = 1$	17	16	16	16
	$p = 2$	26	27	27	27
	$p = \infty$	<b>32</b>	30	30	30
NN		6			
TwoPC		14			

(b)  $n = 100, \delta = 0.04$ 

	$J^*$	4	5	6	7
MULTIPLE	$p = 1$	21	20	17	17
	$p = 2^*$	23	21	20	20
	$p = \infty$	34	32	30	29
PLUGIN	$p = 1$	19	19	19	19
	$p = 2$	32	32	32	32
	$p = \infty$	<b>39</b>	<b>39</b>	<b>39</b>	<b>39</b>
NN		5			
TwoPC		14			

(c)  $n = 400, \delta = 0.02$ 

TABLE 9

Power of the tests for three models of  $(H_1^a)$  with values of  $\delta$  and sample size varying so that  $\sqrt{nd}(f, g)$  remains constant. It appears that those particular sequences of powers are generally non decreasing with the sample size.



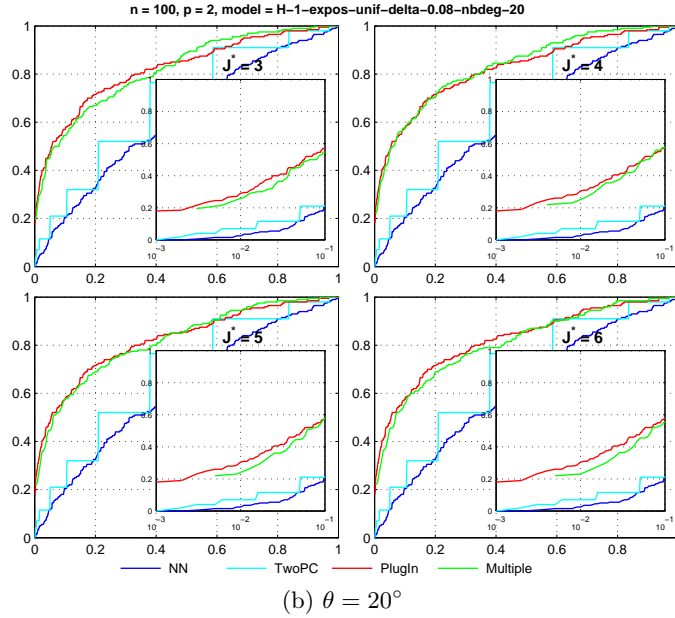
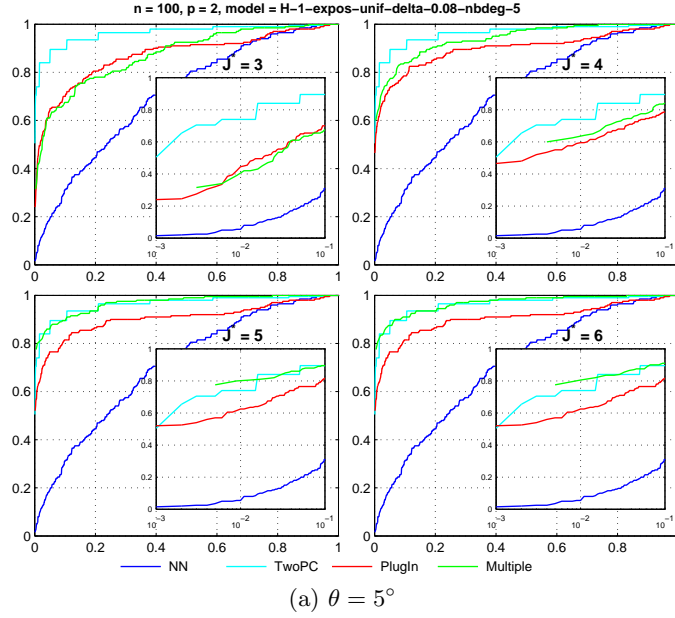


Fig 7: ROC curves for the four different methods and the model ( $H_1^a$ ) with  $n = 100$ ,  $\delta = 8\%$ , and two values of  $\theta$ . For the needlet methods, the  $L^2$  norm is used. It is illustrated here that the behavior of the needlet based tests do not vary a lot with respect to  $\theta$ , contrary to the TwoPC method. See also the text for details and Figure 9. A ROC plots the powers (or true positive rate) in ordinate against the test level (or false positive rate) in abscissa. Insets display the same curves as in the main plot with a logarithmic scale in abscissas, to highlight the comparative performances for relevant level values.

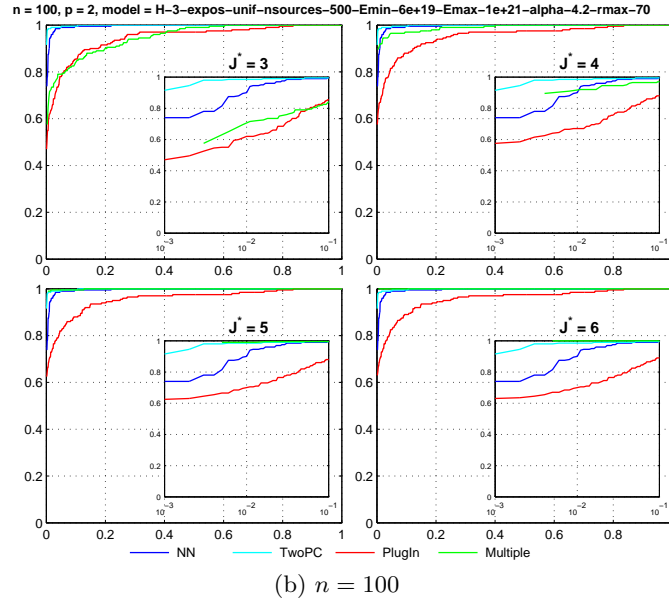
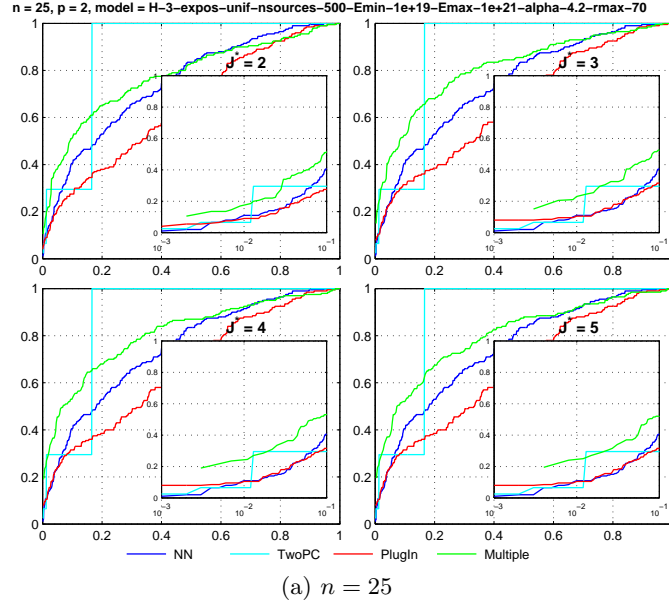


Fig 8: ROC curves for the four methods. For the needlet methods, the de-biased  $L^2$  norm is used. A ROC plots the powers (or true positive rate) in ordinate against the test level (or false positive rate) in abscissa. Insets display the same curves as in the main plot with a logarithmic scale in abscissas, to highlight the comparative performances for relevant level values.

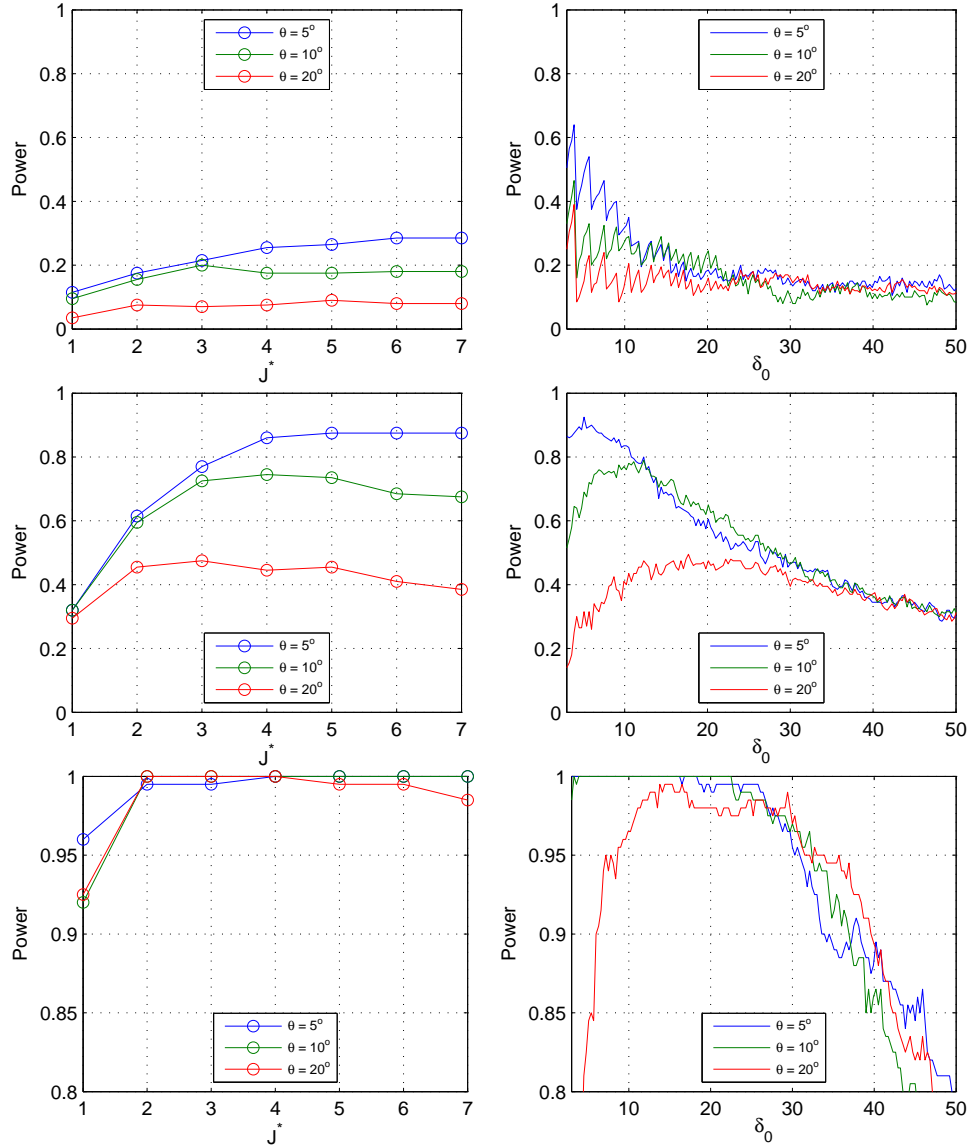


Fig 9: The empirical power associated with the MULTIPLE (left column) and TwoPC (right column) procedures at prescribed levels  $\alpha = 5\%$  with respect to their key parameters  $J^*$  and  $\delta_0$ , respectively. The prescribed levels of the tests are 5%. The three models under consideration are provided by the alternative ( $H_1^a$ ) with  $\theta = 5^\circ, 10^\circ$  and  $20^\circ$ . The number of observations is  $n = 25$  on the first line,  $n = 100$  on the second line,  $n = 400$  on the third line.

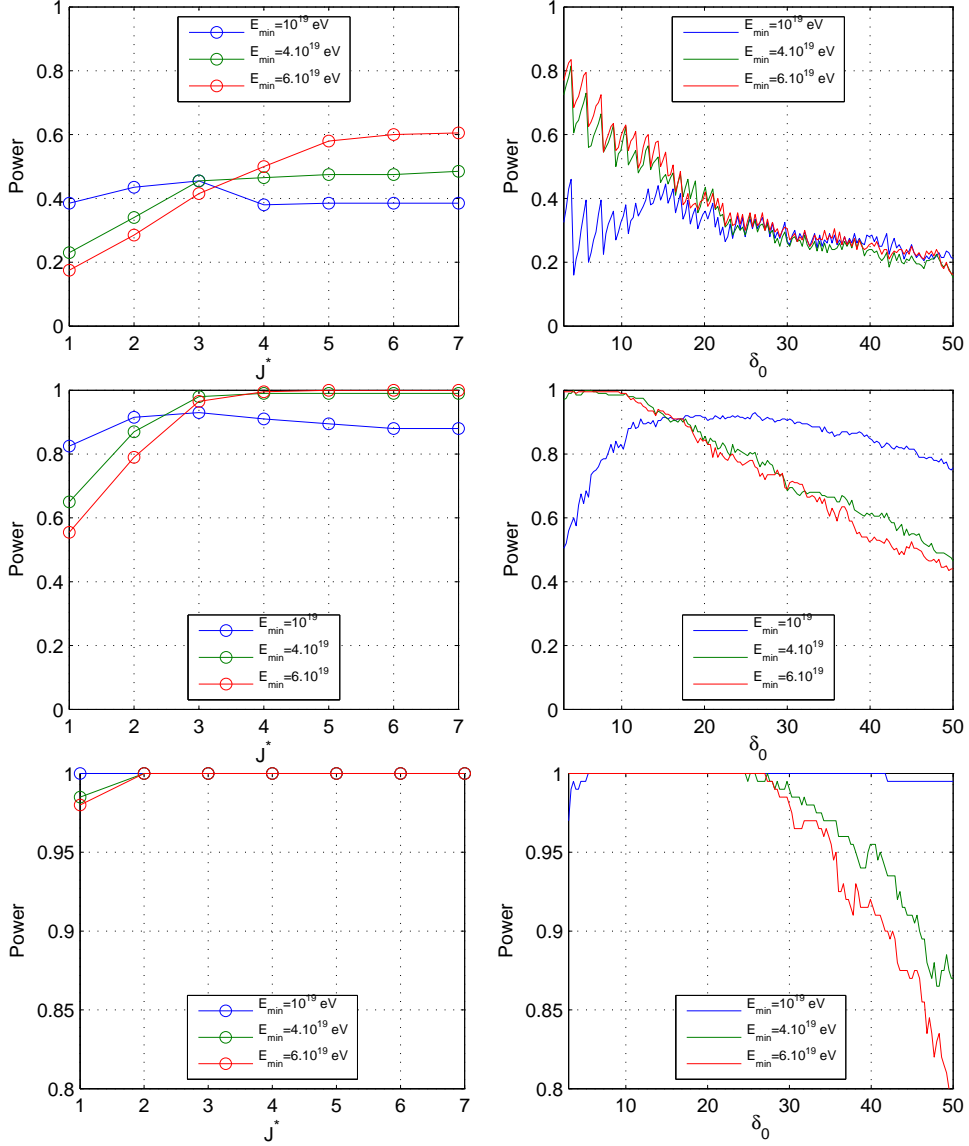


Fig 10: The empirical power associated with the MULTIPLE (left column) and TwoPC (right column) procedures at prescribed levels  $\alpha = 5\%$ , with respect to their key parameters  $J^*$  and  $\delta_0$ , respectively. The three models under consideration are provided by the alternative ( $H_1^c$ ) with  $n_s = 500$ , and  $E_{\min} = 10^{19}$ ,  $4.10^{19}$  or  $6.10^{19}$  eV. The number of observations is  $n = 25$  on the first line,  $n = 100$  on the second line,  $n = 400$  on the third line.

## APPENDIX A: ON-LINE MATERIAL: NEEDLET CONSTRUCTION AND PRACTICAL USAGE

### A.1. Construction of the needlets on the sphere in a nutshell.

A *frame* is a collection of functions with properties close to those of a basis. *Tight frames* share many properties with orthonormal bases, though they are redundant (see Daubechies, 1992, for details).

Needlets were introduced by Narcowich et al. (2006b) as a particular construction of a wavelet tight frame on the sphere. They have been studied in a statistical context ( e.g. Baldi et al., 2006; Baldi et al., 2007) and have also been used recently for cosmological data analysis problems (e.g. ( e.g. Pietrobon et al., 2006; Faÿ et al., 2008; Delabrouille et al., 2009), where they apply to measurements of continuous random fields. An innovative aspect of the present work is to apply them in a detection context for point process data. The most distinctive property of the needlets is their simultaneous perfect localization in the spherical harmonic domain (actually they are spherical polynomials, see Definition 1 below) and potentially excellent localization in the spatial domain.

We recall briefly the construction due to Narcowich et al. (2006b), which is based on three fundamental steps: Littlewood-Paley decomposition (25), splitting (27) and discretization (29). Further details may be found in Baldi et al. (2008), and a discussion on non-tight frame construction in Guillaoux et al. (2009).

Let  $0 \leq a \leq 1$ , be a  $C^\infty$  non negative function defined on  $[0, \infty)$ . Take  $\mathcal{B} > 1$ . We impose  $a$  to be identically 1 on  $[0, 1/\mathcal{B}]$  and compactly supported on  $[0, 1]$ . The function  $b(x) \stackrel{\text{def.}}{=} a(x/\mathcal{B}) - a(x)$  is then a very smooth function supported by  $[1/\mathcal{B}, \mathcal{B}]$ .

Denote by  $\mathcal{H} = L^2(\mathbb{S}, \mu)$  the Hilbert space of square-summable functions on the sphere relatively to the Lebesgue measure  $\mu$  with total mass  $4\pi$ . Let  $\{Y_{\ell m}\}_{\ell \in \mathbb{N}, |m| \leq \ell}$  be the usual spherical harmonics that form an orthonormal basis of  $\mathcal{H}$ . Define  $\mathcal{H}_\ell = \text{span}\{Y_{\ell m}\}_{m=-\ell, \dots, \ell}$  and  $\Pi_\ell$  the orthogonal projector on  $\mathcal{H}_\ell$ . We have

$$\forall f \in \mathcal{H}, \Pi_\ell f = \int_{\mathbb{S}} f(y) L_\ell(\bullet, y) dy \quad (24)$$

with  $L_\ell(x, y) \stackrel{\text{def.}}{=} L_\ell(\text{acos}(x \cdot y))$  and  $L_\ell(\bullet)$  is the Legendre polynomial of order  $\ell$  normalized by the relation  $L_\ell(1) = (2\ell + 1)/(4\pi)$ . Define  $A_0(f) = \frac{1}{4\pi} \int_{\mathbb{S}} f(x) dx$  and the sequence of linear operators

$$B_j \stackrel{\text{def.}}{=} \sum_{\mathcal{B}^{j-1} < \ell < \mathcal{B}^{j+1}} b(\mathcal{B}^{-j}\ell) \Pi_\ell \quad , \quad j \geq 1$$

Obviously,

$$\lim_{J \rightarrow \infty} \|(A_0 + \sum_{j=0}^J B_j)(f) - f\|_2 = 0 \quad (25)$$

Defining

$$B_j(x, y) := \sum_{\mathcal{B}^{j-1} < \ell < \mathcal{B}^{j+1}} b(\mathcal{B}^{-j}\ell) L_\ell(x, y)$$

and

$$D_j(x, y) := \sum_{\mathcal{B}^{j-1} < \ell < \mathcal{B}^{j+1}} \sqrt{b}(\mathcal{B}^{-j}\ell) L_\ell(x, y)$$

we check, using (24) that

$$\forall (x, y) \in \mathbb{S}^2, \quad \int D_j(x, u) D_j(u, y) \, du = B_j(x, y) . \quad (26)$$

$$\forall f \in \mathcal{H}, \quad B_j(f) = \int_{\mathbb{S}} B_j(\bullet, y) f(y) \, dy . \quad (27)$$

Let us now suppose, as important ingredient of the construction of the frame, that there is a *positive quadrature formula* for  $\bigoplus_{l \leq \mathcal{B}^{2+j}} \mathcal{H}_l$ ,  $j \in \mathbb{N}$  which is valid on the sphere (see [Narcowich et al., 2006a,b](#), for an existence result). This means that there exists a finite set  $\mathcal{X}_j = \{\xi_{jk}\}_{k \in \mathcal{K}_j}$  of  $\mathbb{S}$ , and for all  $\xi_{jk}$ , there is an associated coefficient  $\lambda_{jk} > 0$ , such that for all  $f \in \bigoplus_{l \leq \mathcal{B}^{2+j}} \mathcal{H}_l$ , we have the following interpolation formula:

$$\int_{\mathbb{S}} f(y) \, dy = \sum_{k \in \mathcal{K}_j} \lambda_{jk} f(\xi_{jk}).$$

As a consequence we rewrite (26) as

$$B_j(x, y) = \sum_{k \in \mathcal{K}_j} \lambda_{jk} D_j(x, \xi_{jk}) D_j(\xi_{jk}, y) \quad (28)$$

which will directly induce the definition of the needlets. From (27) and (28), we get

$$B_j(f) = \sum_{k \in \mathcal{K}_j} \sqrt{\lambda_{jk}} D_j(\bullet, \xi_k) \int f(y) \sqrt{\lambda_{jk}} D_j(y, \xi_{jk}) \, dy \quad (29)$$

with as a consequence, the following Definition and Proposition.

DEFINITION 1. We define the analysis needlets as follows: for  $k \in \mathcal{K}_j$

$$\psi_{jk}(x) = \sqrt{\lambda_{jk}} D_j(x, \xi_{jk}) = \sqrt{\lambda_{jk}} \sum_{\mathcal{B}^{j-1} < \ell < \mathcal{B}^{j+1}} \sqrt{b}(\mathcal{B}^{-j} \ell) L_\ell(x, \xi_{jk}).$$

PROPOSITION. The set  $\{\psi_{-1}\} \cup \{\psi_{jk}; j \geq 0, k \in \mathcal{K}_j\}$  with  $\psi_{-1,0} \equiv (4\pi)^{-1/2}$  and  $\mathcal{K}_{-1} = \{0\}$  is a frame: for any  $f \in \mathcal{H}$ ,

$$\begin{aligned} \|f\|^2 &= \sum_{j \geq -1, k \in \mathcal{K}_j} |\langle f, \psi_{jk} \rangle|^2 \\ f &= A_0(f) + \sum_j B_j f = \sum_{j \geq -1, k \in \mathcal{K}_j} \langle f, \psi_{jk} \rangle \psi_{jk} \end{aligned}$$

## A.2. Numerical issues.

A.2.1. *Computation of needlet coefficients and related quantities.* The direct computation of the needlet coefficients  $\hat{\beta}_{jk}$  using expression (10) or  $\gamma_{jk}$  using (31) is too lengthy at high scales (in the case of a relatively high number of observations  $n$ ). The evaluations and simple or double sums is not prohibitive for the typical size of  $n$  (100-1000) that we use in our applications but the number of needlet coefficients grows as  $O(\mathcal{B}^{2j})$  with the scale index  $j$ . However, with arbitrary precision, those calculations are advantageously performed in the multipole domain, using existing fast direct and inverse spherical harmonic transform associated to some pixelizations schemes. This restricts the choice of the pixelization.

Namely, from the collection of points  $(X_i)_{i=1, \dots, n}$  and some tessellation of the sphere, one computes an integer valued map  $\tilde{X}$  which counts the number of events falling in each pixel, *i.e.* for each pixel  $p$ ,

$$\tilde{X}_p = \sum_{i=1}^n \mathbf{1}_{X_i \in p}.$$

If the resolution of the pixelization is high enough, this discretization operation on incidence angles can be considered negligible when evaluating the overall performances of the statistical procedures. In practice, we take a discretization finer by far from the minimal requirement. We have used the HEALPix<sup>4</sup> package (Górski et al., 2005), which provides a particular scheme of discretization of the sphere, the associated fast direct and inverse spherical harmonic transforms and utilities. It is very commonly used by cosmic microwave background scientists. In the HEALPix framework, we took

<sup>4</sup><http://healpix.jpl.nasa.gov>

`nside` = 256 meaning that each sky map contains  $12 \times \text{nside}^2 = 786432$  pixels. We denote those points by  $(\xi_p)_{p \in \mathcal{P}}$ . Then, we rely on a fast algorithm to compute the “multipoles”

$$\tilde{a}_{\ell m} \stackrel{\text{def.}}{=} \sum_{p \in \mathcal{P}} \lambda_{jk} Y_{\ell m}(\xi_p) \tilde{X}(\xi_p).$$

The coefficients  $\hat{\beta}_{jk}$  are obtained by smoothing those multipoles by  $\sqrt{b}(2^{-j}\ell)$ , *i.e.* the filter shape at scale  $j$ , and evaluating the corresponding field at the cubature points of scale  $j$  namely

$$\hat{\beta}_{jk} = \lambda_{jk}^{1/2} \sum_{\ell m} \sqrt{b}(2^{-j}\ell) \tilde{a}_{\ell m} Y_{\ell m}(\xi_{jk}).$$

The following diagram sums up those operations

$$(X_i)_{i=1,\dots,n} \rightarrow \text{map } \tilde{X}_p \xrightarrow{\text{SHT}} \tilde{a}_{\ell m} \xrightarrow{\times} \sqrt{b}(2^{-j}\ell) \tilde{a}_{\ell m} \xrightarrow{\text{SHT}^{-1}} \lambda_{jk}^{-1/2} \hat{\beta}_{jk}$$

Double arrows denote as many transforms as the number of scales. On each map of (rescaled) coefficient  $\lambda_{jk}^{-1/2} \hat{\beta}_{j,k}$ , one can apply some thresholding. Then the same filtering operation (multiplication of the multipole moments by the filter  $\sqrt{b}(2^{-j}\ell)$ ) of the maps of coefficients leads to the map  $\sum_k \hat{\beta}_{j,k} \psi_{j,k}$ . Summing those maps over the scales up to  $j = J$  results in the smooth estimate  $\hat{f}_J$ . For details on the practical use and design of the needlet transform of continuous random fields, see [Guiloux et al. \(2009\)](#).

REMARK. Those rather lengthy operations are necessary for accounting a thresholding procedure on the coefficients, as in the case of the `PLUGIN` test, in order to obtain an estimate such as (15). In the `MULTIPLE` approach however, this numerical implementation of the computation of the smooth and linear estimates  $\hat{f}_J$  defined in (11) can be obtained much more directly from the application of a smooth low-pass filter to the multipoles moments of the map  $\tilde{X}_p$ . The transfer function of this low-pass filter is given by the function  $\sum_{j \leq J} b(2^{-j}\ell)$ , which is equal to one up to  $\ell = 2^J$  and then decays to zero smoothly between  $\ell = 2^J$  and  $\ell = 2^{J+1}$ .

As building-block function  $a$ , we have considered here a spline function of order 15 satisfying the above-mentioned conditions (see Figure 11). The corresponding filters  $\ell \mapsto \sqrt{b}(2^{-j}\ell)$  are plotted in the multipole domain. Figure 12 represents the profiles  $\theta \mapsto \sum_{\ell} \sqrt{b}(2^{-j}\ell) L_{\ell}(\cos \theta)$  which are the exact shapes of the needlets at each scale, up to the  $\sqrt{\lambda_{jk}}$  factor. It illustrates the increasing spatial concentration of the needlets as the scale  $j$  grows.



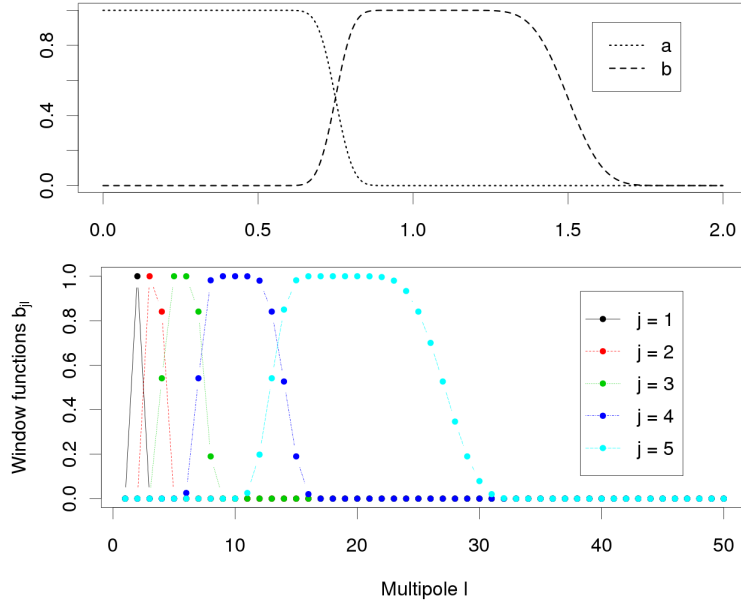


Fig 11: Upper panel: The spline function  $a$  (order 15) and the function  $b = a(\cdot/\mathcal{B}) - a(\cdot)$ . Here  $\mathcal{B} = 2$  so that  $a$  is in  $(0, 1)$  on  $[1/2, 1]$  and  $b$  is compactly supported by  $[1/2, 2]$ . Bottom panel: The filter shapes associated to the five first needlets in multipole domain, namely  $\ell \mapsto \sqrt{b}(2^{-j}\ell)$ ,  $j = 1, \dots, 5$

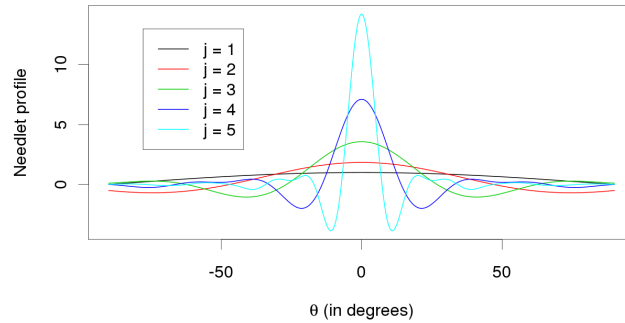


Fig 12: The shape of the five first needlets in the spatial domain as the function of the co-latitude  $\theta$ . Recall that all the  $\psi_{j,k}$  functions are axisymmetric around the points  $\xi_{j,k}$ .

In the `PLUGIN` procedure, the needlet coefficients are thresholded using the data-driven rule (15). The computation of  $\hat{\sigma}_{jk}^2$  and  $\delta_{jk}$  requires additionally to convolve the map  $\tilde{X}$  with the functions  $\psi_{jk}^2$ . This operation is performed along the same lines as above, replacing the factors  $\sqrt{b}(2^{-j}\ell)$  by  $b'_{j,\ell}$  such that

$$\psi_{jk}^2(x) = \lambda_{jk} \sum_{\ell < 2^{j+2}} b'_{j,\ell} L_\ell(x, \xi_{jk}).$$

**A.2.2. Choice of the needlets.** Relations between the regularity of  $a$  and the asymptotic localization properties of the needlets is described in [Narcowich et al. \(2006b\)](#) (see also [Guilloux et al., 2009](#)). Here, because of the smoothness of the exposure function  $g$  of the instrument, the localization of the needlet is not such a sensitive issue as in the case of analysis of a random field with missing data. We have chosen dyadic needlets ( $\mathcal{B} = 2$ ) and  $a$  being a spline function of order 15, which leads to simple but sufficiently concentrated analysis wavelets. Those needlets are represented in Figure 11 and 12. Note however that asymptotic localization properties are theoretical properties and that in our small sample studies, maximal scales involves angular frequencies smaller than  $\ell = 128$ . This is not “high frequency” compared to usual analysis in CMB analysis for instance.

**A.2.3. Computation of  $L^p$  distances.** Let  $\hat{f}$  be some estimate of  $f$  based on the coefficients  $\hat{\beta}_{j,k}$ . Whatever adaptive needlet method we use, we need to measure the discrepancy between the observations and the null hypothesis using  $L^p$  norms with  $p = 1, 2, \infty$ , say. An exact numerical evaluation of an  $L^p$  distances between  $\hat{f}$  and  $g$  requires some quadrature formula. Given a high resolution scheme and for any  $p \geq 1$ , we approximate  $L^p$  norms and distances by estimating  $\hat{f}$  and  $g$  on a regular grid (or pixelization)  $(\xi_k)_{k=1,\dots,N}$ , and then by computing

$$\hat{d}_p(f, g) \stackrel{\text{def.}}{=} d_p(\hat{f}, g) = \|\hat{f} - g\|_p \simeq \left( \sum_{k=1,\dots,N} \lambda_k |\hat{f}(\xi_k) - g(\xi_k)|^p \right)^{1/p} \quad (30)$$

for some quadrature weights  $(\lambda_k)$ . We have used above-mentioned Healpix weights at resolution `nside=256`. In some particular cases ( $p = 2$  for instance), an exact evaluation is possible. Assume for example that both  $\hat{f}$  and  $g$  are “band-limited” in the sense that they belong to  $\text{span}\{\psi_{jk}, j \leq J, k \in \mathcal{K}_j\}$ . Then, thanks to the properties of the needlets (tight frame, semi-orthogonality)

$$\|\hat{f} - g\|_2^2 = \sum_{j \leq J+1} \sum_{k \in \mathcal{K}_j} (\beta_{jk}(\hat{f}) - \beta_{jk}(g))^2.$$

However, with  $\beta_{jk}(\hat{f}) = \hat{\beta}_{jk}$  (defined in Eq. 10),  $(\hat{\beta}_{jk}(f) - \beta_{jk}(g))^2$  is a biased estimate of  $(\beta_{jk}(f) - \beta_{jk}(g))^2$  whereas an unbiased estimate is provided by

$$\zeta_{jk} \stackrel{\text{def.}}{=} \frac{1}{n(n-1)} \sum_{i \neq i'} (\psi_{jk}(X_i) - \beta_{jk}(g))(\psi_{jk}(X_{i'}) - \beta_{jk}(g)). \quad (31)$$

Thus,  $\|f - g\|_2^2$  may be estimated unbiasedly by

$$\hat{d}_{2^*}(f, g) = \sum_{j \leq J+1} \sum_{k \in \mathcal{K}_j} \zeta_{jk}. \quad (32)$$

Note that the  $\zeta_{jk}$ 's and this estimate are not necessarily positive. Moreover, this technique does not extend to non linear estimates of  $f$  that use thresholded coefficients.

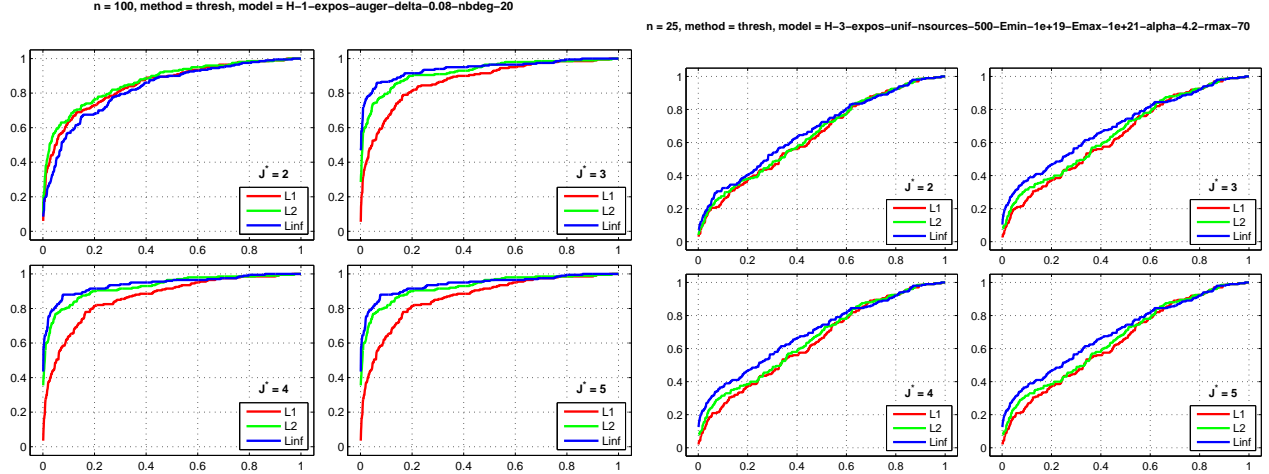


Fig 13: Effect of the choice of the norm on the sensitivity of the MULTIPLE procedure, depending on the shape of the density under the alternative. Top: Unimodal density given by model ( $H_1^a$ ) with  $\delta = 0.08$ ,  $\theta = 20^\circ$  (sample size  $n = 100$ ). Bottom: Multimodal density given by model ( $H_1^c$ ) with  $E_{\min} = 10^{19}$  eV,  $n_s = 500$  (sample size  $n = 25$ ).

GILLES FAÏ  
LABORATOIRE MAS  
ECOLE CENTRALE PARIS  
GRANDE VOIE DES VIGNES  
92 295 CHÂTENAY-MALABRY, FRANCE  
E-MAIL: [gilles.fay@ecp.fr](mailto:gilles.fay@ecp.fr)

JACQUES DELABROUILLE  
LABORATOIRE APC, CNRS UMR7164  
10 RUE A. DOMON ET L. DUQUET  
75013 PARIS, FRANCE  
E-MAIL: [DELABROUILLE@APC.UNIV-PARIS7.FR](mailto:DELABROUILLE@APC.UNIV-PARIS7.FR)

GÉRARD KERKYACHARIAN  
UNIVERSITÉ PARIS X-NANTERRE  
CNRS LPMA  
175 RUE DU CHEVALERET,  
75013 PARIS, FRANCE  
E-MAIL: KERK@MATH.JUSSIEU.FR

DOMINIQUE PICARD  
UNIVERSITÉ PARIS DIDEROT,  
CNRS LPMA  
175 RUE DU CHEVALERET,  
75013 PARIS, FRANCE  
E-MAIL: PICARD@MATH.JUSSIEU.FR